

1 TCCTCCGGTC GCCC GCCCTC GGGGCAGCTA GTGGCGCAGC CCCCCGCCG
 51 CGGCCCTGGC CTCCCGGGCG GCGCGCAGG GGAGGGGTTA AGCTGCCGA
 101 GGGACCGCCG CGTGCGGGGC GAGAGGGAGC CCCC GGTTGGCAG
 151 CGGGCGGGGT TCGGTCCGAG CCCGGTGGGA GGCTCCCGGA GCGCAGCCTG
 201 GGCCCAGCCC ACCCCGCGCC GGCGGCCATG GCAGGCACCC TGGACCTGGA
 251 CAAGGGCTGC ACGGTGGAGG AGCTGCTCCG CGGGTGCATC GAAGCCTTCG
 301 ATGACTCCGG GAAGGTGGCGG GACCCGCAGC TGGTGCAGCAT ATTCCCTCATG
 351 ATGCACCCCT GGTACATCCC CTCCCTCTAG CTGGCGGCCA AGCTGCTCCA
 401 CATCTACCAA CAATCCCGGA AGGACAACCTC CAATTCCCTG CAGGTGAAA
 451 CGTGCCACCT GGTCAGGTAC TGGATCTCCG CCTTCCCAAGC GGAGTTGAC
 501 TTGAACCCGG AGTTGGCTGA GCAGATCAAG GAGCTGAAGG CTCTGCTAGA
 551 CCAAGAAAGGG AACCGACGGC ACAGCAGCCT AATCGACATA GACAGCGTCC
 601 CTACCTACAA GTGGAAGCGG CAGGTGACTC AGCGGAACCC TGTGGGACAG
 651 AAAAAGCGCA AGATGTCCT GTTGTGAC CACCTGGAGC CCATGGAGCT
 701 GGCGGAGCAT CTCACCTACT TGGAGTATCG CTCCCTCTGC AAGATCCTGT
 751 TTCAGGACTA TCACAGTTTC GTGACTCATG GCTGCAGTGT GGACAACCC
 801 GTCTCGGAGC GGTTCATCTC CCTCTTCAAC AGCGTCTCAC AGTGGGTGCA
 851 GCTCATGATC CTCAGCAAAC CCACAGCCCC GCAGCGGGCC CTGGTCA
 901 CACACTTTGT CCACGTGGCG GAGAACGCTGC TACAGCTGCA GAACTTCAAC
 951 ACGCTGATGG CAGTGGTGGGG GGGCTGAGC CACAGCTCCA TCTCCCGCCT
 1001 CAAGGAGACC CACAGCCACG TTAGCCCTGA GACCATCAAG CTCTGGGAGG
 1051 GTCTCACGGA ACTAGTGACG GCGACAGGCA ACTATGGCAA CTACCGGCGT
 1101 CGGCTGGCAG CCTGTGTGGG CTTCCGCTTC CCGATCCTGG GTGTGCACCT
 1151 CAAGGACCTG GTGGCCCTGC AGCTGGCACT GCCTGACTGG CTGGACCCAG
 1201 CCCGGACCCG GCTCAACGGG GCCAAGATGA AGCAGCTCTT TAGCATCTG
 1251 GAGGAGCTGG CCATGGTGAC CAGCCTGCAGG CCACCAAGTAC AGGCAACCC
 1301 CGACCTGCTG AGCCTGCTCA CGGTGTCCTC GGATCAGTAT CAGACGGAGG
 1351 ATGAGCTGTA CCAGCTGTC CTGCAGCGGG AGCCGCGCTC CAAGTCCTCG
 1401 CCAACCAGCC CCACGAGTTG CACCCACCA CCCC GGCCCCCGG CGGTACTGG
 1451 GGAGTGGACC TCGGCTGCCA AACCCAAGCT GGATCAGGCC CTCGTGGTGG
 1501 AGCACATCGA GAAGATGGTG GAGTCGTGT TCCGGAACTT TGACGTCGAT
 1551 GGGGATGGCC ACATCTACA GGAAGAATTG CAGATCATCC GTGGGAACCTT
 1601 CCCCTTACCTC AGCGCCTTGT GGGACCTCGA CCAGAACCGAG GATGGCTGCA
 1651 TCAGCAGGGGA GGAGATGGTT TCCTATTTC TCGCCTCCAG CTCTGTGTTG
 1701 GGGGGCGCA TGGGCTTCGT ACACAACCTC CAGGAGAGCA ACTCCTTGCG
 1751 CCCCGTCGCC TGCCGCCACT GCAAAAGCCCT GATCCTGGGG ATCTACAAAGC
 1801 AGGGCCTCAA ATGCCGAGCC TGTGGAGTGA ACTGCCACAA GCAGTGCAAG
 1851 GATCGCCTGT CAGTTGAGTG TCGGCCAGG GCCCAGAGTG TGAGCCTGG
 1901 GGGGTCTGCA CCCTCACCC CACCCATGCA CAGCCACCAT CACCGCGCCT
 1951 TCAGCTTCTC TCTGCCCGC CCTGGCAGGC GAGGCTCCAG GCCCTCAGCA
 2001 ATCCCCCTCC CAGCAGAGAT CCGTGAGGAG GAGGTACAGA CGGTGGAGGA
 2051 TGGGTGTTT GACATCCACT TGTAATAGAT GCTGTGGTTG GATCAAGGAC
 2101 TCATTCTGC CTTGGAGAAA ATACTTCAAC CAGACCAAGGG AGCCTGGGG
 2151 TGTCGGGGCA GGAGGCTGGG GATGGGGGTG GGATATGAGG GTGGCATGCA
 2201 GCTGAGGGCA GGGCCAGGGC TGGTGTCCCT AAGGTTGTAC AGACTCTTGT
 2251 GAATATTTGT ATTTTCCAGA TGGAAATAAAAA AGGCCCGTGT AATTAAAAAA
 2301 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA (SEQ ID NO:1)

FEATURES:

5'UTR: 1-227
 Start Codon: 228
 Stop Codon: 2073
 3'UTR: 2076

FIGURE 1, page 1 of 2

Homologous proteins:

Top 10 BLAST Hits

| | | Score | E |
|--------------------|--|-------|-------|
| CRA 1000682340958 | /altid=gi 6358505 /def=gb AAF07219.1 AF043722... | 1293 | 0.0 |
| CRA 18000005086608 | /altid=gi 5031623 /def=ref NP_005816.1 RAS ... | 1241 | 0.0 |
| CRA 18000005188697 | /altid=gi 6755290 /def=ref NP_035372.1 RAS,... | 1202 | 0.0 |
| CRA 18000005205935 | /altid=gi 7662334 /def=ref NP_056191.1 KIAA... | 618 | e-175 |
| CRA 18000005188699 | /altid=gi 3928857 /def=gb AAC79700.1 (AF081... | 533 | e-150 |
| CRA 18000005152782 | /altid=gi 9507035 /def=ref NP_062084.1 RAS ... | 531 | e-149 |
| CRA 18000005192860 | /altid=gi 7242201 /def=ref NP_035376.1 RAS ... | 529 | e-149 |
| CRA 18000005192861 | /altid=gi 4038292 /def=gb AAC97349.1 (AF106... | 526 | e-148 |
| CRA 18000005188698 | /altid=gi 5032025 /def=ref NP_005730.1 RAS ... | 525 | e-148 |
| CRA 1000733831533 | /altid=gi 6650545 /def=gb AAF21898.1 AF081197... | 525 | e-148 |

BLAST dbEST hits:

| | Score | E |
|---|-------|-------|
| gi 5432583 /dataset=dbest /taxon=9606 ... | 1310 | 0.0 |
| gi 9876673 /dataset=dbest /taxon=960... | 1281 | 0.0 |
| gi 11286864 /dataset=dbest /taxon=96... | 1249 | 0.0 |
| gi 11285315 /dataset=dbest /taxon=96... | 1207 | 0.0 |
| gi 5432584 /dataset=dbest /taxon=9606 ... | 733 | 0.0 |
| gi 4372300 /dataset=dbest /taxon=9606 ... | 720 | 0.0 |
| gi 12295751 /dataset=dbest /taxon=96... | 700 | 0.0 |
| gi 12288965 /dataset=dbest /taxon=96... | 599 | e-168 |
| gi 6920402 /dataset=dbest /taxon=960... | 573 | e-161 |
| gi 2005039 /dataset=dbest /taxon=9606 ... | 573 | e-161 |

EXPRESSION INFORMATION FOR MODULATORY USE:

library source:

From BLAST dbEST hits:

gi|5432583 Testis
gi|9876673 Liver-non-cancerous
gi|11286864 Brain glioblastoma
gi|11285315 Brain glioblastoma
gi|5432584 Testis
gi|4372300 B Cell Chronic lymphatic leukemia
gi|12295751 Adult marrow
gi|12288965 Adult marrow
gi|6920402 Lymph germinal center B cell
gi|2005039 Lymph

From tissue screening panels:

Leukocyte

1 MAGTILDLDKG CTVEELLRGK IEAFDDSGKV RDPQLVRIFL MMHPWYIPSS
51 QLAAKLLHIY QOSRKDNSNS LQVKTCFLVR YWISAFPAEF DLNPELAEQI
101 KELKALLDQE GNRRHSSLID IDSVPTYKWK RQVTQRNPVG QKKRKMSLIF
151 DHLEPMELAE HITYLEYRSF CKILFQDYHS FVTHGCTVDN PVLERFISLF
201 NSVSQWVQLM ILSKPTAPQR ALVITHFVHV AEKLLQLQNF NTLMAVVGG
251 SHSSISRLKE THSHVSPETI KLWEGLTEL V ATGNYGNR RRLAACVGF
301 FPILGVHLKD LVALQLALPD WLDPARTRLN GAKMKQLFSI LEELAMVTS
351 RPPVQANPDL LSLLTVSLDQ YQTEDELYQL SLQREPRSKS SPTSPTSCP
401 PPRPPVLEEW TSAAKPKLDQ ALVVEHIEKM VESVFRNFDV DGDGHISQEE
451 FQIIRGNFPY LSAFGDLDQN QDGCIISREEM VSYFLRSSV LGGRMGFVHN
501 FQESNSLRPV ACRHCKALIL GIYKQGLKCR ACGVNCHKQC KDRLSVECRR
551 RAQSVSLEGS APSPSPMHSH HHRAFSFSLP RPGRGSRPP AIPLPAEIRE
601 EEVQTVEDGV FDIHL (SEQ ID NO:2)

FEATURES:

Functional domains and key regions:

[1] PDO00004 PS00004 CAMP_PHOSPHO_SITE

cAMP- and cGMP-dependent protein kinase phosphorylation site

Number of matches: 3

| | | |
|---|---------|------|
| 1 | 113-116 | RRHS |
| 2 | 144-147 | RKMS |
| 3 | 584-587 | RRGS |

[2] PDO00005 PS00005 PKC_PHOSPHO_SITE

Protein kinase C phosphorylation site

Number of matches: 7

| | | |
|---|---------|-----|
| 1 | 27-29 | SGK |
| 2 | 63-65 | SRK |
| 3 | 126-128 | TYK |
| 4 | 134-136 | TQR |
| 5 | 269-271 | TIK |
| 6 | 349-351 | SLR |
| 7 | 506-508 | SLR |

[3] PDO00006 PS00006 CK2_PHOSPHO_SITE

Casein kinase II phosphorylation site

Number of matches: 9

| | | |
|---|---------|------|
| 1 | 12-15 | TVEE |
| 2 | 63-66 | SRKD |
| 3 | 117-120 | SLID |
| 4 | 163-166 | TYLE |
| 5 | 339-342 | SILE |
| 6 | 373-376 | TEDE |
| 7 | 447-450 | SQEE |
| 8 | 476-479 | SREE |
| 9 | 605-608 | TVED |

[4] PDO00008 PS00008 MYRISTYL

N-myristoylation site

Number of matches: 4

| | | |
|---|---------|--------|
| 1 | 19-24 | GCIEAF |
| 2 | 249-254 | GLSHSS |
| 3 | 284-289 | GNYGNY |
| 4 | 492-497 | GGRMGF |

[5] PDOC00009 PS00009 AMIDATION
Amidation site

582-585 PGRR

[6] PDOC00018 PS00018 EF HAND
EF-hand calcium-binding domain

Number of matches: 2
1 439-451 DVGDGHISQEEF
2 468-480 DQNQDGCIISREEM

[7] PDOC00379 PS00479 DAG_PE_BIND_DOM_1
Phorbol esters / diacylglycerol binding domain

499-548 HNFQESNSLRPVACRHCKALILGIYKQGLKCRACGVNCHKQCKDRLSVEC

Membrane spanning structure and domains:

| Helix | Begin | End | Score | Certainty |
|-------|-------|-----|-------|-----------|
| 1 | 34 | 54 | 0.713 | Putative |
| 2 | 195 | 215 | 0.653 | Putative |
| 3 | 238 | 258 | 0.788 | Putative |

BLAST Alignment to Top Hit:

>CRA|18000005086608 /altid=gi|5031623 /def=ref|NP_005816.1| RAS guanyl
releasing protein 2 (calcium and DAG-regulated); calcium
and diacylglycerol-regulated guanine nucleotide exchange
factor I [Homo sapiens] /org=Homo sapiens /taxon=9606
/dataset=nraa /length=609
Length = 609
Score = 1241 bits (3176), Expect = 0.0
Identities = 608/615 (98%), Positives = 609/615 (98%)
Frame = +3

Query: 228 MAGTLDDKGCTVEELLRCIEAFDDSGKVRDPQLVRIFLMMHPWYIPSSQLAAKLLHIY 407
MAGTLDDKGCTVEELLRCIEAFDDSGKVRDPQLVR+FLMMHPWYIPSSQLAAKLLHIY
Sbjct: 1 MAGTLDDKGCTVEELLRCIEAFDDSGKVRDPQLVRMFLMMHPWYIPSSQLAAKLLHIY 60

Query: 408 QQSRSKDNNSNLSQVKTCHLVRYWISAFPAEFDLNPELAEQIKEALKALLDQEGRNRHSSLID 587
QQSRSKDNNSNLSQVKTCHLVRYWISAFPAEFDLNPELAEQIKEALKALLDQEGRNRHSSLID
Sbjct: 61 QQSRSKDNNSNLSQVKTCHLVRYWISAFPAEFDLNPELAEQIKEALKALLDQEGRNRHSSLID 120

Query: 588 IDSVPPTYKWKRQVTQRNPVGQKKRKMSLLFDHLEPMELAEHLTYLEYRSFCKILFQDYHS 767
IDSVPPTYKWKRQVTQRNPVGQKKRKMSLLFDHLEPMELAEHLTYLEYRSFCKILFQDYHS
Sbjct: 121 IDSVPPTYKWKRQVTQRNPVGQKKRKMSLLFDHLEPMELAEHLTYLEYRSFCKILFQDYHS 180

Query: 768 FVTHGCTVDNPVLERFISLFNSVSQWVQLMILSKPTAPQRALVITHFVVAEKLLQLQN 947
FVTHGCTVDNPVLERFISLFNSVSQWVQLMILSKPTAPQRALVITHFVVAEKLLQLQN
Sbjct: 181 FVTHGCTVDNPVLERFISLFNSVSQWVQLMILSKPTAPQRALVITHFVVAEKLLQLQN 240

Query: 948 NTLMAVVGGLSHSSISRLKETHSHVSPETIKLWEGLTELVTATGNYGNYRRRLAACVGF 1127
NTLMAVVGGLSHSSISRLKETHSHVSPETIKLWEGLTELVTATGNYGNYRRRLAACVGF
Sbjct: 241 NTLMAVVGGLSHSSISRLKETHSHVSPETIKLWEGLTELVTATGNYGNYRRRLAACVGF 300

Query: 1128 FPILGVHLKDLVALQI ALPDWLDPARTRLNGAKMKQKFSLILEELAMVTSRPPVQANPDL 1307
FPILGVHLKDLVALQI ALPDWLDPARTRLNGAKMKQKFSLILEELAMVTSRPPVQANPDL
Sbjct: 301 FPILGVHLKDLVALQI ALPDWLDPARTRLNGAKMKQKFSLILEELAMVTSRPPVQANPDL 360

Query: 1308 LSLLTVSLDQYQTEDELYQLSLQREPRSKSSPTSPTCTPPRPPVLEEWTSAAKPKLDQ 1487
LSLLTVSLDQYQTEDELYQLSLQREPRSKSSPTSPTCTPPRPPVLEEWTSAAKPKLDQ
Sbjct: 361 LSLLTVSLDQYQTEDELYQLSLQREPRSKSSPTSPTCTPPRPPVLEEWTSAAKPKLDQ 420

Query: 1488 ALVVEHIEKMVESVFRNFDVGDGHISQEEFQIIRGNFPYLSAFGDLQDGCISREEM 1667
ALVVEHIEKMVESVFRNFDVGDGHISQEEFQIIRGNFPYLSAFGDLQDGCISREEM
Sbjct: 421 ALVVEHIEKMVESVFRNFDVGDGHISQEEFQIIRGNFPYLSAFGDLQDGCISREEM 480

Query: 1668 VSYFLRSSSVLGGRMGFVHNFQESNSLRPVACRHCKALILGIYKQGLKCRACGVNCHKQC 1847
VSYFLRSSSVLGGRMGFVHNFQESNSLRPVACRHCKALILGIYKQGLKCRACGVNCHKQC
Sbjct: 481 VSYFLRSSSVLGGRMGFVHNFQESNSLRPVACRHCKALILGIYKQGLKCRACGVNCHKQC 540

Query: 1848 KDRLSVECRRRAQSVSLEGSAPS PSMHSHHRAFSFSLPRPGRRGSRPPAIPLPAEIRE 2027
KDRLSVECRRRAQSVSLEGSAPS PSMHSHHRAFSFSLPRPGRRGSRPP EIRE
Sbjct: 541 KDRLSVECRRRAQSVSLEGSAPS PSMHSHHRAFSFSLPRPGRRGSRPP-----EIRE 594

Query: 2028 EEVQTVEDGVFDIHL 2072
EEVQTVEDGVFDIHL
Sbjct: 595 EEVQTVEDGVFDIHL 609 (SEQ ID NO:4)

>CRA|1000682340958 /altid=gi|6358505 /def=gb|AAF07219.1|AF043722_1
(AF043722) guanine exchange factor MCG7 isoform 1 [Homo
sapiens] /org=Homo sapiens /taxon=9606 /dataset=nraa
/length=671
Length = 671
Score = 1293 bits (3309), Expect = 0.0
Identities = 639/670 (95%), Positives = 643/670 (95%), Gaps = 4/670 (0%)
Frame = +3

| | | |
|---|---|------|
| Query: 75 | GRGGVKLPGQPPRAGREGAPGGGAAG----GVRSEPGRLPERSLGPAPAPAAMAGTL | 242 |
| GRG P + +E G +G | GVRSEPGRLPERSLGPAPAPAAMAGTL | |
| Sbjct: 8 | GRGTQGWPGSSEQHQEATSSAGLHSGVDELGVRSEPGRLPERSLGPAPAPAAMAGTL | 67 |
| Query: 243 | DLDKGCTVEELLRCIEAFDDSGKVRDPQLVRIFLMMHPWYIPSSSQLAAKLLHIYQQSRK | 422 |
| DLDKGCTVEELLRCIEAFDDSGKVRDPQLVR+FLMMHPWYIPSSSQLAAKLLHIYQQSRK | | |
| Sbjct: 68 | DLDKGCTVEELLRCIEAFDDSGKVRDPQLVRMFLMMHPWYIPSSSQLAAKLLHIYQQSRK | 127 |
| Query: 423 | DNSNSLQVKTCHLVRYWISAFPAEFDLNPELAEQIKEKALLDQEGRNRRHSSLIDIDSVP | 602 |
| DNSNSLQVKTCHLVRYWISAFPAEFDLNPELAEQIKEKALLDQEGRNRRHSSLIDIDSVP | | |
| Sbjct: 128 | DNSNSLQVKTCHLVRYWISAFPAEFDLNPELAEQIKEKALLDQEGRNRRHSSLIDIDSVP | 187 |
| Query: 603 | TYWKWKRQVTQRNPVGQKKRKMSLLFDLEPMELAEHLTYLEYRSFCKILFQDYHSFVTHG | 782 |
| TYWKWKRQVTQRNPVGQKKRKMSLLFDLEPMELAEHLTYLEYRSFCKILFQDYHSFVTHG | | |
| Sbjct: 188 | TYWKWKRQVTQRNPVGQKKRKMSLLFDLEPMELAEHLTYLEYRSFCKILFQDYHSFVTHG | 247 |
| Query: 783 | CTVDNPVLERFISLNFNSVSQWVQLMILSKPTAPQRALVITHFVHVAEKLLQLQNFNTLMA | 962 |
| CTVDNPVLERFISLNFNSVSQWVQLMILSKPTAPQRALVITHFVHVAEKLLQLQNFNTLMA | | |
| Sbjct: 248 | CTVDNPVLERFISLNFNSVSQWVQLMILSKPTAPQRALVITHFVHVAEKLLQLQNFNTLMA | 307 |
| Query: 963 | VVGGLSHSSISRLKETHSHVSPETIKLWEGLTELVTATGNYGNYRRRLAACVGFRFPILG | 1142 |
| VVGGLSHSSISRLKETHSHVSPETIKLWEGLTELVTATGNYGNYRRRLAACVGFRFPILG | | |
| Sbjct: 308 | VVGGLSHSSISRLKETHSHVSPETIKLWEGLTELVTATGNYGNYRRRLAACVGFRFPILG | 367 |
| Query: 1143 | VHLKDLVALQLALPDWLDPARTRLNGAKMKQLFSILEELAMVTSLRPPVQANPDLLSLLT | 1322 |
| VHLKDLVALQLALPDWLDPARTRLNGAKMKQLFSILEELAMVTSLRPPVQANPDLLSLLT | | |
| Sbjct: 368 | VHLKDLVALQLALPDWLDPARTRLNGAKMKQLFSILEELAMVTSLRPPVQANPDLLSLLT | 427 |
| Query: 1323 | VSLDQYQTEDELYQLSLQREPRSKSSPTSPSCTPPRPPVLEEWTSAAKPKLDQALVVE | 1502 |
| VSLDQYQTEDELYQLSLQREPRSKSSPTSPSCTPPRPPVLEEWTSAAKPKLDQALVVE | | |
| Sbjct: 428 | VSLDQYQTEDELYQLSLQREPRSKSSPTSPSCTPPRPPVLEEWTSAAKPKLDQALVVE | 487 |
| Query: 1503 | HIEKMVESVFRNFDVGDGHISQEEFQIIRGNFPYLSAFGDLQDGCISREEMVSYFL | 1682 |
| HIEKMVESVFRNFDVGDGHISQEEFQIIRGNFPYLSAFGDLQDGCISREEMVSYFL | | |
| Sbjct: 488 | HIEKMVESVFRNFDVGDGHISQEEFQIIRGNFPYLSAFGDLQDGCISREEMVSYFL | 547 |
| Query: 1683 | RSSSVLGGRMGFVHNFQESNSLRPVACRHCKALILGIYKQGLKCRACGVNCHKQCKDRRLS | 1862 |
| RSSSVLGGRMGFVHNFQESNSLRPVACRHCKALILGIYKQGLKCRACGVNCHKQCKDRRLS | | |
| Sbjct: 548 | RSSSVLGGRMGFVHNFQESNSLRPVACRHCKALILGIYKQGLKCRACGVNCHKQCKDRRLS | 607 |
| Query: 1863 | VECRRAQSLEGSAPS PSMHSHHRAFSFSLPRPGRRGSRPPAIPLP AEIREEEVQT | 2042 |
| VECRRAQSLEGSAPS PSMHSHHRAFSFSLPRPGRRGSRPP | EIREEEVQT | |
| Sbjct: 608 | VECRRAQSLEGSAPS PSMHSHHRAFSFSLPRPGRRGSRPP-----EIREEEVQT | 661 |
| Query: 2043 | VEDGVFDIHL 2072 | |
| VEDGVFDIHL | | |
| Sbjct: 662 | VEDGVFDIHL 671 (SEQ ID NO:5) | |

FIGURE 2, page 4 of 7

>CRA|18000005188697 /altid=gi|6755290 /def=ref|NP_035372.1| RAS,
guanyl releasing protein 2; RAP 1A protein-specific
guanine nucleotide exchange factor 1; CalDAG-GEFI [Mus
musculus] /org=Mus musculus /taxon=10090 /dataset=nraa
/length=608
Length = 608
Score = 1202 bits (3076), Expect = 0.0
Identities = 589/615 (95%), Positives = 597/615 (96%)
Frame = +3

Query: 228 MAGTLDLKGCTVEELLRCIEAFDDSGKVRDPQLVRIFLMMHPWYIPSSQLAAKLLHIY 407
MA TLDLKGCTVEELLRCIEAFDDSGKVRDPQLVR+FLMMHPWYIPSSQLA+KLLH Y
Sbjct: 1 MASTLDLKGCTVEELLRCIEAFDDSGKVRDPQLVRFMFLMMHPWYIPSSQLASKLLHFY 60

Query: 408 QQSRKDNSNSLQVKTCHLVRYWISAFPAEFDLNPELAEQIKELKALLDQEGRNRRHSSLID 587
QQSRKDNSNSLQVKTCHLVRYW+SAFPAEFDLNPELAE IKELKALLDQEGRNRRHSSLID
Sbjct: 61 QQSRKDNSNSLQVKTCHLVRYWVSAFPAEFDLNPELAEPIKELKALLDQEGRNRRHSSLID 120

Query: 588 IDSVPPTYWKWRQVTQRNPVGOKKRKMSLLFDHLEPMELAEHLTYLEYRSFCKILFQDYHS 767
I+SVPTYWKWRQVTQRNPV QKKRKMSLLFDHLEPMELAEHLTYLEYRSFCKILFQDYHS
Sbjct: 121 IESVPPTYWKWRQVTQRNPVEQKKRKMSLLFDHLEPMELAEHLTYLEYRSFCKILFQDYHS 180

Query: 768 FVTHGCTVDNPVLERFISLFNSVSQWVQLMILSKPTAPQRALVITHFVVAEKLLQLQNF 947
FVTHGCTVDNPVLERFISLFNSVSQWVQLMILSKPTA QRALVITHFVVAEKLLQLQNF
Sbjct: 181 FVTHGCTVDNPVLERFISLFNSVSQWVQLMILSKPTATQRALVITHFVVAEKLLQLQNF 240

Query: 948 NTLMAVVGGLSHSSISRLKETHSHVSPETIKLWEGLTELVTATGNYGNRRLAACVGFR 1127
NTLMAVVGGLSHSSISRLKETHSHVSP+TIKLWEGLTELVTATGNY NYRRLAACVGFR
Sbjct: 241 NTLMAVVGGLSHSSISRLKETHSHVSPDTIKLWEGLTELVTATGNSNYRRLAACVGFR 300

Query: 1128 FPILGVLKDLVALQLALPDWLDPARTRLNGAKMKQLFSILEELAMVTSLRPPVQANPDL 1307
FPILGVLKDLVALQLALPDWLDP RTRLNGAKM+QLFSILEELAMVTSLRPPVQANPDL
Sbjct: 301 FPILGVLKDLVALQLALPDWLDPGRTRLNGAKMRQLFSILEELAMVTSLRPPVQANPDL 360

Query: 1308 LSLLTVSLDQYQTEDELYQLSLQREPRSKSSPTSPSCTPPRPPVLEEWTSAAKPKLDQ 1487
LSLLTVSLDQYQTEDELYQLSLQREPRSKSSPTSPSCTPPRPPVLEEWTS AKPKLDQ
Sbjct: 361 LSLLTVSLDQYQTEDELYQLSLQREPRSKSSPTSPSCTPPRPPVLEEWTSVAKPKLDQ 420

Query: 1488 ALVVEHIEKMVESVFRNFDVGDGHISQEEFQIIRGNFPYLSAFGDLQDQDGCIISREEM 1667
ALV EHIKEKMVESVFRNFDVGDGHISQEEFQIIRGNFPYLSAFGDLQDQDGCIISREEM
Sbjct: 421 ALVAEHIEKMVESVFRNFDVGDGHISQEEFQIIRGNFPYLSAFGDLQDQDGCIISREEM 480

Query: 1668 VSYFLRSSSVLGGRMGFVHNFQESNSLRPVACRHCKALILGIYKQGLKCRACGVNCHKQC 1847
+SYFLRSSSVLGGRMGFVHNFQESNSLRPVACRHCKALILGIYKQGLKCRACGVNCHKQC
Sbjct: 481 ISYFLRSSSVLGGRMGFVHNFQESNSLRPVACRHCKALILGIYKQGLKCRACGVNCHKQC 540

Query: 1848 KDRLSVECRRRAQSVSLEGSAPS PSMHSHHRAFSFSLPRPGRRGSRPPAIPLPAEIRE 2027
KDRLSVECRRRAQSVSLEGSAPS PSP H+ HHRAFSFSLPRPGRR SRPP EIRE
Sbjct: 541 KDRLSVECRRRAQSVSLEGSAPS PPTHT-HHRAFSFSLPRPGRRSSRPP-----EIRE 593

Query: 2028 EEVQTVEDGVFDIHL 2072
EEVQTVEDGVFDIHL
Sbjct: 594 EEVQTVEDGVFDIHL 608 (SEQ ID NO:6)

>CRA|18000005205935 /altid=gi|7662334 /def=ref|NP_056191.1| KIAA0846
protein [Homo sapiens] /org=Homo sapiens /taxon=9606
/dataset=nraa /length=689
Length = 689
Score = 618 bits (1576), Expect = e-175
Identities = 314/597 (52%), Positives = 409/597 (67%), Gaps = 5/597 (0%)
Frame = +3

Query: 234 GTLDDDKGCTVEELLRCIEAFDDSGKVRDPQLVRIFLMMHPWYIPSSQLAAKLLHIYQQ 413
G+ L K T++ELL CIE FDD+G++ + L RI L+MH WY+ S++LA KLL +Y+
Sbjct: 2 GSSGLGKAATLDELLCTCIEMFDDNGE LDNSYLPRLVLMHRWYLSSTELAEKLLCMYRN 61

Query: 414 SRKD NSNLSLQVKTCHLVRYWISAFPAEFDLNPELA EQI KELK ALLD QEG NRRHSSLIDID 593
+ ++ N ++K C+ +RYWI FPAEF+L+ L +E + + Q G +H SLIDI
Sbjct: 62 ATGESCN EFRLKIC YFMRYWILKFP AEFNLDLGLIR MTEEFREVAS QOLGYEKH VSLIDIS 121

Query: 594 SVPTYKWK RQVTQRN PVGQKKR KMSL LFHDHLE P MELA EH LT YLE R SFCK I LFQDYHSFV 773
S+P+Y W R+VTQR V KK K LLFDHLE P+ELAEH LT+LE++SF +I F DY S+V
Sbjct: 122 SIPS YDW MRR VTQR KKVS -KKG KAC LLF DHL EPIELA EH LT FLE HKS FRR ISFT DY QSYV 180

Query: 774 THGCTVDNPVLERFISLFNSV SQWVQ LMILSKPTAPQ RALV ITHF VHVA EKLL QLQ NFNT 953
HGC +NP LER I+LFN +S+W VQ LM+LSKPT QRA VIT F++VA+KLL QL+N FNT
Sbjct: 181 IHGC LENNPTLERSIALFNGISKWVQ LMVL SKPTP Q QRAEV ITKF INVAK LLQ LKNFNT 240

Query: 954 LMAVVG GLSH SSI SRLKETH SHVSPETIKLWE GLTEL VTA TGN YGN YR RRLAAC VGF RFP 1133
LMAVVG GLSH SSI SRLKETH SH+S E K W +TEL V++ GNY NYR+ A C GF+ P
Sbjct: 241 LMAVVG GLSH SSI SRLKETH SHL SEVT K NW NEMTEL VSS NG NCYR KA FAD CDG FKIP 300

Query: 1134 ILGVHLKDLVALQ LALPDWL DPART RLNGAKM KQLFSI LEELAM VTS LRPPVQANPDLLS 1313
ILGVHLKDL+A+ + PDW + ++N KM QL L EL + + ++ N DL++
Sbjct: 301 ILGVHLKDLIAVHV IFPDW TE--ENKVNI VKM HQLS VT LSEL VSLQ NASH HLE PNMDL IN 358

Query: 1314 LLTVSL DQYQ TEDE LYQ LSLQ REPR SKSS P TSCT P P R P P VLE EWT SAAKPKL DQ AL 1493
LLT+SLD Y TED++Y+LSL EPR+ SPT SPT+ P +P V EW PK D +
Sbjct: 359 LLT LSL DLYH TEDDI YK LSLV LEP RNS KSP TSPTT ---PNKP VV PLE ALGV MPK P DPTV 415

Query: 1494 VVEHIEK MVESVFRN FDVG DGHIS QEEF QI IRGNF PYLSA FG DLDQ NQDG C ISREEM VS 1673
+ + HI K+ VESV FRN+ D D DG+ ISQE+F+ I NFP+ L +F LD++ QDG IS++ EM++
Sbjct: 416 INKH IRKL VESV FRN YDH DHDG YI SQED FEST IAAN FPFL DSCV LDQ DGL ISK DEMMA 475

Query: 1674 YFLRSSV VL GGRM --GFVHN FQESNSL RPVACRH CKALI LGIYKQ GLK CRAC GVNC HKQC 1847
YFLR+ S L +M GF+ HNF QE L+P C HC + GI KQG KC+ CG NCH KQC
Sbjct: 476 YFLRAKSQLHCKMGPFIHNFQEMTYLKPTFC EHCAGFLWGI IKQGYKCKDCGANCH KQC 535

Query: 1848 KDRLS VECRR RAQS VSL---EGSAPSPSPM HS HH RAFS FS L P R P G RGS RPPAI PL 2009
KD L + CRR A++ SL GS P + F F G R AI L
Sbjct: 536 KDLLV LAC RR FARAP SLS SGH G SLP GSP S L PPAQ D E V F E P G V TAGH RD L DS RAI TL 592 (SEQ ID NO:7)

>CRA|18000005188699 /altid=gi|3928857 /def=gb|AAC79700.1| (AF081196)
calcium and DAG-regulated guanine nucleotide exchange
factor II [Rattus norvegicus] /org=Rattus norvegicus
/taxon=10116 /dataset=nraa /length=795
Length = 795
Score = 533 bits (1358), Expect = e-150
Identities = 267/590 (45%), Positives = 390/590 (65%), Gaps = 12/590 (2%)
Frame = +3

Query: 156 GVRSEPGG RL PER SLG PAHPAPAAMAGTL D-----LDKGCTVEELLRCIEAFDD S 308
G R+ P GRL +S PA ++A L KG ++++L+ CI++FD
Sbjct: 17 GSRAGPKGR LEAK STNSPLPAQPSLAQITQFRMMVSLGH LAKG ASL DLD LIDSCI QSF DAD 76

Query: 309 GKV-RDPQLVRIFLMMHPWYIPSSQLAAKLLHIYQQSRKDNSNSLQVKTCHLVRWISAF 485
 G + R QL+++ L MH I S++L KL+++Y+ + + NS + +K C+ VRYWI+ F
 Sbjct: 77 GNLCRSNQLLQVMLTMHRIIISSAELLQKLMNLYKDALEKNSPGICLKICYFVRYWITEF 136

 Query: 486 PAEFIDLNELAEQIKEKLALLDQEGRNRRHSSLIDIDSVPPTYKWKRQVTQRNPVG-QKKRK 662
 F ++ L ++E + L+ G H LID + + W R++TQR KKRK
 Sbjct: 137 WIMFKMDASLTSTMEEFQDLVKANGEESHCHLIDTTQINSRDWSRKLQTQRIKSNTSKRK 196

 Query: 663 MSLLFDHLEPMELAEHLTYLEYRSFCKILFQDYHSFVTHGCTVDNPVLERFISLFNSVSQ 842
 +SLLFDHLEP EL+EHLTYLE++SF +I F DY +++ + C +NP +ER I+L N +SQ
 Sbjct: 197 VSLLFDHLEPEELSEHLTYLEFKSFRRISFSQNYLVNSCVKENPTMERSIALCNGISQ 256

 Query: 843 WVQLMILSKPTAPQRALVITHFVVAEKLLQLQNFNTLMAVVGLSHSSISRLKETHSHV 1022
 WVQLM+LS+PT RA V F+HVA+KL QLQNFNTLMAV+GGL HSSISRLKET SHV
 Sbjct: 257 WVQLMVLSPRTPQLRAEVFIKFIHVAQKLHQLQNFNTLMAVIGGLCHSSISRLKETSSHV 316

 Query: 1023 SPETIKLWEGLTTELVTATGNYGNYRRRLAACVGFRFPILGVHLKDLVALQALPDWLDPA 1202
 E K+ +TEL+++ NY NYRR C F+ PILGVHLKDL++L A+PD+L+
 Sbjct: 317 PHEINKVLGEMTELLSSCRNYDNYRAYGETHFKIPILGVHLKDLISLYEAMPDYLEDG 376

 Query: 1203 RTRLNGAKMKQLFSILEELAMVTSRPPVQANPDLLSLLTVSLDQYQTEDELYQLSLQRE 1382
 ++N K+ L++ + EL + + PP+ AN DL+ LLT+SLD Y TEDE+Y+LS RE
 Sbjct: 377 --KVNVOKLALYNHINELVQLQDVAPPDANKDLVHLLTLSLDLYTEDEIYELSYARE 434

 Query: 1383 PRSKSSPTSPSTSCTPPPRPPVLEEWTSAAKPKLDQALVVEHIEKMVESVFRNFDVDGDGH 1562
 PR+ +P P +PPV+ +W S PK D + +H+++MV+SVF+N+D+D DG+
 Sbjct: 435 PRNRHAPP----LTPSKPPVVWDWASGVSPKDPKTI SKHVQRMVDSVFKNYLDQDGY 489

 Query: 1563 ISQEEFQIIRGNFPYLSAFGDLQDQNQDGCIISREEMVSYFLRSSSVLGG-RMGFVHNQES 1739
 ISQEEF+ I +FP+ +F +D+++G ISR+E+ +YF+R+SS+ +GF HNFQE+
 Sbjct: 490 ISQEEFEKIAASFP--SFCVMDKDREGLISRDEITAYFMRASSIYSKLGFLGFPHNFQET 547

 Query: 1740 NSLRPVACRHCKALILGIYKQGLKCRACGVNCHKQCKDRLSVECRRRAQS 1889
 L+P C +C + G+ KQG +C+ CG+NCHKQCKD + EC++R++S
 Sbjct: 548 TYLKPTFCDCNCA GFLWGVVIKQGYRCKDCGMNCHKQCKDLVVFECKKRSKS 597 (SEQ ID NO:8)

Hummer search results (Pfam):

| Model | Description | Score | E-value | N |
|---------|--|-------|---------|---|
| PF00617 | RasGEF domain | 123.5 | 4e-33 | 1 |
| PF00130 | Phorbol esters/diacylglycerol binding domain | 59.5 | 3.6e-14 | 1 |
| PF00036 | EF hand | 21.8 | 0.00027 | 2 |
| PF01237 | Oxysterol-binding protein | 3.5 | 4.2 | 1 |

Parsed for domains:

| Model | Domain | seq-f | seq-t | hmm-f | hmm-t | score | E-value |
|---------|--------|-------|--------|-------|--------------|---------|---------|
| PF01237 | 1/1 | 249 | 272 .. | 1 | 24 [. 3.5 | 4.2 | |
| PF00617 | 1/1 | 148 | 336 .. | 1 | 227 [] 123.5 | 4e-33 | |
| PF00036 | 1/2 | 430 | 458 .. | 1 | 29 [] 17.4 | 0.0047 | |
| PF00036 | 2/2 | 463 | 482 .. | 5 | 24 .. 6.7 | 4.9 | |
| PF00130 | 1/1 | 499 | 548 .. | 1 | 51 [] 59.5 | 3.6e-14 | |

1 ACAGAAAGGT CCTGTTCTA AGTCTTACAT TACCAAGACT GAGGTGCAGG
 51 GGCGGTCTCG GATCCCCCGC CCCAAGGCTG GGAGGGGCAC GCCTCGGAAG
 101 GGAGGTTGGG GGTGGTGGT TTACACAGTGA GTGTGTCTGA AGCCAAATGG
 151 TCGGAAACCG TTACCCGCTC TCCTAGGCC CCGTAGTGGG GACCCCAACC
 201 GCCTGCGGCT GCCCTCTCCA AGTTCTCCC TGTTGGCCAG GCATCCAGGT
 251 CTCCAGTCTC CGAGCTGGG AGAACCCACC GCCACATGCG GCTGCCCTT
 301 TCCATTGAC CCTGTGGGA GCCAGGCTTC CGGGGCCCG TTCCCTCTGT
 351 GTGAACCTGGG CCCCCCGCCC CCATTCCCAG ACATCAAGGC CGCGTCTCCA
 401 GATAGCCACG ATTTCAITCC TCGCTCCCCA CAGGTCCCTC TCCCCAAAT
 451 ATTCCCATCT TGTCTAGCC CATCCCCAG ACTATCTCAA GGACCAGCTG
 501 TCCCCACGCC CCCGACCTCC ACTAGGCCCTG TGCCACCCCG TGCCCTGCAGG
 551 AAGACGCCCG GTCCCGGGCC GGGTTAGCCC CATGGGAACCT GTTTGTCTCG
 601 AAAACAGGAA CCCGAGCTGG GGGCTGGGC GGGCGCCCTC TCCCCACCGC
 651 AGTCCGCTTC CTGCCCCCTCC CCGCTTCCCTC CGCCCGACAC CCAGGCAGGG
 701 CGGGGGGCAC TGGGGCTCC CGGGTTGGG GAGGGGCTCT TCGTTTCCGGT
 751 CCCCCCTCCC CGCTCCCGGG CGGGGGGCC TCCGGTCCGC CGCCTCGGGG
 801 CAGCTAGTGG CGCAGCCCCC CGCCCGCGGC CCTGGCCTCC CGGGCGGC
 851 GGCAGGGAGG GGGTTAAGCT GCGCAGGGG CGCCCGCGTG CGGGCGAGA
 901 GGGAGCCCCC GGTGGGGGTG GCGCAGCCGG CGGTGCGGAG CTCCCGCAG
 951 GGGCGGAGGG GGGAGGGGGC AGCCTGGC GGGGGCGGGG GCGGGCGGC
 1001 GGGGAGCGGG GCCGCGGGCT GGAGAGCGGG CGGGAGGCC AGCCGCAGG
 1051 AGGGCGGCCG GCGGGAGCGC ACGGAGGTGG CGTCGGCCAG GCGGGTGC
 1101 GCTCCCTGCG GCAGGTCCCA AGAGTGACTG GCGGAGCGC GCGGGGGCGC
 1151 CAGGCAGAGG AGGGCGGCC CCCAGCGAC TCCCCCCCCC CCCAGGGCGG
 1201 CGCGGGCGGG CTGGGGCGGG CGAGGGGTG GGGAGTCTGC GGCGGGGTC
 1251 TGGGAGAGGG GGCAGCGGCC ACAGAGAGCTA AGGCGCGCTG GATCCCCGGA
 1301 GGGCGGAGGA CCTCCACGGT GCACCCAGCT TTTCCAGCC ACCTCCAGC
 1351 GGGGCCCTCC CCCCGTACC CCCATTGGC AGATGAGAAA ATTGAGGCTC
 1401 CCAGAGGCCA AGTGTATTCTC AAGGTACAC GAGGAAGCGG TAGAGCCAGG
 1451 CGGGGACGCC TCTGGGTGGC TCTTAGGAAA AGTCCGCTG AGAAACTCCGT
 1501 ACAGGAGCTC CCTGTCTCC CAGCCTGGG GAGTGAGTAT GTTAGGGCC
 1551 GGGGTACCTT TCCGTGGGC AAGGCTCTGC CAAAATCTGG GAGTGAGGG
 1601 AGTCAGGGAG CTGGGGCCGC AGGGCGGGCC CTGCACCCGA AATGGGAGGG
 1651 GGGCGACGGA ATGGGCGTGC GCACCCATGG GGGTGTGTGC ATGTGTGTGG
 1701 GAGTGTACAT CGCTGGAGAG GCACTGCCCT GCGTGTGTGC ACACGTGTGA
 1751 GGATGTCAGC GCCTGTGTGG CGCGGGGACT CAAGGCTGGG CTGGCTCAAG
 1801 TGAACAGCAC GTCCAGGAGG CGACCTCGTC CGGGGGTTTG CATTCTGGGG
 1851 TGGACGAGCT GGGTATGTG GCCTGAGGGT TTCTTCGTGC AGGTGTGCAC
 1901 AGGGTGTGGG TGCCATTGTG TGTGAGAGAC GGAGGATGGG GAGGCCGGTG
 1951 CCTGTGGCC GGTGCGTGA AGTGCGGACG CCTGCACCTC CACTTAGGTC
 2001 CCCGGCCTCC GACGACTAAC TTGGGTGTGG AGTGTGGGCC CCTGCCAGGG
 2051 TGGGTATGAC CCCGCCAGTG ACCGGAGTTG CTAATGGTGT CATGCCACCA
 2101 CCGGCCACCC TTGGCGCGAG CGCCCCCCTC TGACACCCCT GCTCCGTGCG
 2151 CGCTCACAGT TCGCCTGTGC GGGGCCGGGG CAAGGGTCAG GAGCCGGGGA
 2201 TAGGGAGGAA GAGGGCCTGT GGACAAGCTG AGCCGGGACC CCTGGGACCT
 2251 TTGCGGAGGT GGCCTGGGAG CGCTCAGTTC CCAGGCTGAG GCTTCCCGCT
 2301 GACGCCCTCT GGCGCGAGCG GGCTCCCCCC GCCCCAGGAA TGTTCTCTC
 2351 CCATCCAGTC CGCCTCCCCCT AGGGCAGGCC CCCTGGGGC TGCCGAGCC
 2401 CGGCCTCGCC TTCCCTGGCT CCCGGGAGGG GGCAGGGCAGA GCAGGACGCC
 2451 TGGTTCTCT CCCCCCACCT CCCATACCAAG GGAGAAATTC CTCCGAGGTC
 2501 CCCTCAGGCT CTGGGTTCCC AAAATAACCC TGCGGGGGAA GGGAGGCTGT
 2551 GGAGGGAGGG AAGCGGGAGG GGCGCAGAGC CGAGCTGCGG GGTGCTGCAG
 2601 GTGCTCTGG GGAGAGGGCG CGAGGAGAAAG GCGCCCTGCG GGGGGCTGGG
 2651 CGCCAGGCC CGCTGGGATC TTGGTTCGTC CCCATCCTCG TGAAGCCCCCT
 2701 CGGGCTTCCC GCGACTCCCAG GGGTGGGCCG GAAGCCTCTC TGCGGGTCCG
 2751 TTTCCCAACT GGCGGGTTGC ACCATCCCCG GCCAGACCGT TTAACCCCGG
 2801 GACTGGCCGC GGGGGACAACT TCCGCCCCCTG TCCAGCAGGG GGCAGCCCC
 2851 CCCCGCCCCG TTTCTGCCCG CGGGGCCGCT CCCCGCCCG CGACTCCGCA
 2901 GACTCCCCTC CTGCCTCTCC CGGGACAGGG GTTCGGTCCG AGCCCGGTGG
 2951 GAGGCTCCCCG GAGCGCAGCC TGGGCCCAGC CCACCCCGCG CGGGCGGCCA
 3001 TGGCAGGCAC CCTGGACCTG GACAAGGGCT GCACGGTGGA GGAGCTGCTC
 3051 CGCGGGTGC ACGAAGCCTT CGGTGAGTGG CTCGGGAGGG CACACGGAGC
 3101 CTGAGCCTAG CCCCGAGTCT GAGCCCCGGT CCCTGCCTCC CAGGCACAGT

FIGURE 3, page 1 of 12

3151 CCAGGGCACA GCCCTGACCC GGACCCACCC TGCTCCGCAG CGTGCAGTCT
 3201 CTTAACGAA AGCCTCCTCC GCAACGCAGG GCAGAGAGAT GCACGCCCTT
 3251 CAGACAGATG AGGTTTCCCT TCTCTAGCCT TCCCAGCGG CGGGGAAGGG
 3301 AGGGCCGGT CCCGGACTCT GACACTTGAG GGGCATTATC TGTCTCCCGG
 3351 GGAATCCGGA GGAACTCGCT ATCTCCGGCC TGGGAGCTGT TTCCGGCTAA
 3401 TGGGGGCGG CTTATCTGGT GAAGGGGTGC CCCCTCCCCC CAAGCGCTCA
 3451 GGAAATGACC TCTGGATTCT TGACCCCGGG GAACCCAGGC TCCCTCCGCC
 3501 CCAGCTGGTT CCCCTCCGGA CGATGGCGG CTGGGGCGCT CCCCTCCCTC
 3551 AGTCCTCAGG GCGTGCCTAT CTCTCGCCCA CCACACCTTT CCTCTCTAAT
 3601 TTGCCTCCTG CTCTCGGAGT CCTGGGAAG CAGGAGGTGG GCGGGGTGGA
 3651 GCGTCACCC GAAGGACCGA TACCTGGGG GTTGGGGGGT GAGGATGAGG
 3701 CATGTAGCT GCGGACCA CTCAGCCACC TGCTTTGAC CCTTCGGAGT
 3751 CAGATGACTC CGGGAAAGGT CGGGACCCGC AGCTGGTGC GATGTTCCCTC
 3801 ATGATGCACC CTGGTACAT CCCCTCCCT CAGCTGGCGG CCAAGCTGCT
 3851 CCACATATAT CTTTCGGG CCTTGCCAAG GCCCCCGCCG TCGGAGCCCA
 3901 TGCGCAGCCC CTCTGCCAG CCCAGGTGCA GAATGAGCCT CGCTCCTAAG
 3951 TATAAGGCCAC TCCTTATCCC AGAGCTCAGG CGTCGTCCCA GCCTCCAAGT
 4001 AGGGCCTAGG CTCTGCCCTC TCCTTGCTCC TAGCGACTCG GTCCGTCCC
 4051 CAGGCTCTGT CCCCAGCCGA GGCCTTGCCC TCCTTCTCCC TAGAGTCTAG
 4101 GGCCTGCCCT TGCTTCAGGC TTGGGTGCGC CCCGTGCATC TCTCTCTCCC
 4151 AGAGCCCAGG CTTTGCTTC AGCCTCCCTC AGCACCTAGT CCTCCACCCC
 4201 CACCTCCAAC CCCTCCCAGA GCTCAAGCCT CACCCCCAGC ATCTCCGCAG
 4251 AGCGCAAGCC CCATCCCTAG AACGTGTCTC CTAGAACAG GCCCCGCC
 4301 CAGCCTCCCT CCACGCAGGC CTCCCTTCT AGAGTTAACG GGCCTCCTTA
 4351 ACCCTCTCCT TCACCTACCA ACAATCCCGG AAGGACAAC CCAATTCCCT
 4401 GCAGGTGAAA ACGTGCCACC TGGTCAGGTG AGTCTTCCC CTGGGGCTCT
 4451 AGCCCTCCCT CTTCTCCCT TCTCTCTGGC TTCAGGCTGG CCTGGAGGAG
 4501 GGGGCAGGGC GCTGTTCTG GGAGTGGGTT TGAACCTGG CTTGTCCGG
 4551 TGGGCAGTGC TGCCACAGGC TCACCCCTTC CTGGGTCTGG GCCTTAATT
 4601 TCTTTCTGC GCAGTGCAGG TGTTGTCTC AAGGGTCAA TGTACACTTG
 4651 GAGTGGCGAA GGAAAGAGCT GGAACCATAG TTTGAGGGTC TTTTGCTTA
 4701 GGTGACTATA ATCTCAAATA GCTCCTGCA GCCTGCTGG TGATGGTGGG
 4751 GGAAGGGCTA TCTTGGGTGA CTCCCCGCTC CTCCAGGTAC TGGATCTCCG
 4801 CCTTCCCAGC GGAGTTTGAC TTGAACCCGG AGTTGGCTGA GCAGATCAAG
 4851 GAGCTGAAGG CTCTGCTAGA CCAAGAAGGG AACCGACGGC ACAGCAGCCT
 4901 AATCGACATA GACAGCGTGT GCCTGGGGGG ACCACAGAGG GCTGGGGGG
 4951 CACTCAGTAT CCTATACCAT CTGTGCTTAA TAAATGTCTG TTGAACTGAA
 5001 TGAGTGAGGG TCATGTTGCT CTCTCGCTTA AAAACCTTCC ATGGCTCCCT
 5051 ATTGCCTTCA ACATGCCCTC TCTGGCAGC TTGGCGTTCC TGCCTCATCT
 5101 TCCACTGCCA CCACCCATCC CACACACCTC CTCCGTAGC TGCCTGGGT
 5151 CGGCTCCCCG TCGGCTGAGC TCTCGAGTCC TTTCTCATCA TGGTGCTCTG
 5201 CTCATATCAT CCCCCCTTGGT GCCTCCCTCG TTGTTACCAAG ACTCAGTTCA
 5251 GGCATGAAGT CTCCGTGGC TCTGAGGGTT CGGGGCTCTT CGGGGGTAGA
 5301 ATTGTCGTT CCCACCTCTG TTTTCATGG CACTTTGTAC AGACTCCTGT
 5351 ACAAAAGACCT CTGTACATGT GTCACCGCTGT TTTGTGATCA TGTGTTTCTG
 5401 TGTCTGTCTC CCTCAGTAGA CTGTGAGCTC CTCGAGGGCA GGAACCGTGT
 5451 CTTACTCATC TCTGTATTCC CAGCGCCTAG CACAGTGCCT GGCACAGAGT
 5501 ACGTTGTTCA TAAATGTGTG TTGAGTGCAT GACGGGGTGG GGGGAGATGA
 5551 GGAGGAGTTG CTGGGACTGG GAACATTCTG GCCTAGGACA GTGCCTCGCA
 5601 TTATGTAGGT TCTCAGTAAG CGTGAATGGT GTGTCTGTGT GAGTGGGGGG
 5651 CCACGAGGCA TGCGCATGTC CAGCAAAGGG CTCACTACCC CTGCCCCCCC
 5701 AGCCCTACCT ACAAGTGGAA GCGGCAGGTG ACTCAGCGGA ACCCTGTGGG
 5751 ACAGAAAAAG CGCAAGATGT CCCTGTTGTT TGACCCACCTG GAGCCCATGG
 5801 AGCTGGCGGA GCATCTCAC TACTTGGAGT ATCGCTCCCT CTGCAAGATC
 5851 CTGGTGCAGC CCGAGGGCTG GGGGGTCAGG GGTCCAATGT GGGCTGGAAG
 5901 AGAGTTCTAG GAGGGGAGG GTCCCTGGCG TAGGCTGGGT CACAGGGTGC
 5951 ATCAGGGGTT TCAGTGTAAAC CACTGAAGGT CAGCTGGAGG GTGAGGAGTG
 6001 GCTATCAGTG AGGGGAGAGG CCGGCAAGGT GCTGAGGCCA CTCCCTCATGC
 6051 CCCCAGTTTC AGGACTATCA CAGTTTCGTT ACTCATGGCT GCACTGTGGA
 6101 CAACCCCGTC CTGGAGCGGT TCATCTCCCT CTTCACACAGC GTCTCACAGT
 6151 GGGTGCAGCT CATGATCCCTC AGCAAACCCA CAGCCCCGCA GCGGGCCCTG
 6201 GTCATCACAC ACTTTGTCCA CGTGGCGGGAG GTGCCTGCC CTCCTCCCG
 6251 GTGTCTCCCA ACCACCCAC ATGCCAGTCA GGCAACCCCT TCCCTCCCC

FIGURE 3, page 2 of 12

6301 TAACCCACTG CCTTCTCTCT AGATAAGCTG GGCAAATTC TGGGCCACT
6351 CAGTGACTCC CTGCCTCTCC GTCCCCATT GCCTTCCAGA AGCTGCTACA
6401 GCTCGAGAAC TTCAACACGC TGATGGCAGT GGTCGGGGC CTGAGCCACA
6451 GCTCCATCTC CCGCCTCAAG GAGACCCACA GCCACGTTAG CCCTGAGACC
6501 ATCAAGGTGC CTGGGACTGG GGAGGGGCG GTGCTTCCA GGTCTGTCTT
6551 CACTGGTCC TCCCAGCAGC ACTGGGGCT GGGCACAGCT GTCCCATTT
6601 GATAGATATG GAAATGGAGG CTCAGAGGG TTAAGTGCTT TTCTCAGTTT
6651 GCACAATGGC AACAGCAGAG TGGGGCTCA CAGGTCGTCA GGGACCCCAA
6701 AGCTAGTACT TTTTTTTTTT TTTTTAAGAC AGGGTCTCTC TCTCTGTGT
6751 CCAGACTGGA GTTCAGTGGT GCAGTCACAA GCTCACTGCA GCCTTGAATT
6801 CCTGAGCTCA ATCGATCTC CCACCTCAGC CTCTGAGTA GCTGGGACTA
6851 CAGGTGTACG CCACATGCC TAATTTTGT ATTGTTATTA ATTTTTTTT
6901 TTTTTTTTA GAGATGGGT TTTGCATGT TGCCAGACT GGTCTTGAAC
6951 TCCTGGGCTC AAGTATCG CCTGCCTTGG CCTCCCAAAG TGCTGAGATT
7001 ATGGCTTGAG CCATTGTCG TTGCCATTG TAGTTCTTC TTTTCTTCT
7051 CCTTCATTTT TTATTATTT TGAAGTATTG TGAAGTATTG AGTAACATAC
7101 ATATAGAAAA GTATATAAAA ACATATGAGA CTGGCGTAG TAGTCACAC
7151 CTGTAATCCC AGCACTTTGG GAGGCTGAGG TGGCAGATC ACGTGACATC
7201 AGGAGTTTGA GACCAGCTG GCAACAAAGG TGGAAACCCA TCTCTACTAA
7251 AATACAAAAA TTAGCCAGGC ATGGTGGCAC GCACCTGGAA TCCAAGCTAC
7301 TTGGGAGGCT GAGGCAGGAG GAGAATTACT TGAACTCAGG AGGCAGGAGT
7351 TGCACTGAGC CAAGATTGTG CCACCTCACT CCAGCCTGGG CGACAGAGTG
7401 AGACTCCATC TAAAAAAAAA GAAAAGTATA TAAAAACATA TGAATAGTTT
7451 AAAGAAAAAT TGTAAGAAA ACACTGTGTA ACTACTGCC GGGTTGGGAA
7501 ATAGAACCTT GCCAGGCCCC CAAGCGCCCA GCACCTTAA GAATAACTCC
7551 CTCCCCACGA CTTTGCAAT GATGATCTT CTTTCTTTA TAGCTTCACC
7601 ATGTAGGTAT GCGGTCCAAA ACAATGTGGG GCTTTTGTT GTCTGTTTG
7651 AACTTCTAT GAATGGAATG TTGTTGTGT TATTTATGT CTTGCTTTT
7701 TCATTCCACA TGTTCTGAG AGTCTTTCA TTCTGTATG TGGAGCAATT
7751 GTTTTTCAT TTTCAATTGCC ATATAATATT TTATTGTACG TCTACCCCAA
7801 TTCATTATT TATTTATTTT TTTGAGATGG AGTCTGTCTC TGTCACTCAG
7851 GCTGGAGTGC GGTGGCGAGA TCTCATCACT GCAACTTCCG TCTCCTGGT
7901 TTACGTGATT CTCGTGCCTC AGCCCTCTGA GTAGCTGGGA TTATGGGCTC
7951 GTACCACAC GTCTGGCTAA TTTTTGTAG AGACAGGCTT TCACCATGTT
8001 GCCGAGGCTG GTCTTGAAC CCTGAGCTCA GGCAATCCAC CCGCTTTAGC
8051 CTCCCAAAGT GCTGGGATTA CAGGTGTGAG CCACGCC CAGCTTACCC
8101 CAATTATGT ATTGATTCTA TTGTTGAATG TTGGGGTTTT TCCTTTCTT
8151 TTCTTCTTT CTTTTCTTT CTTTTTCTT TTGTTGGAG GAGGGAGTCT
8201 TGCTCTGTG CCAAGCTGGA GTGCAGTGAC GCTAATTGGG CTCACTGCA
8251 CACTGCACCC TCTGCCTCCC GGGTCAAGC GATTCTCTG CCTCAGCCTC
8301 CTGAGTAGCT GGGACTACAG GCATGCACCA CCACACCGG CTAATTGGT
8351 TATTTTTTA GTAGAGATGA GTTCCACC ATGTTGCCA AGATGGTCTC
8401 CATCTCTGAA CCTCATGATC CATCTGCAT GGCCTCCCAA AGTGTGAGA
8451 TTACAAGTGT GAGCCACAC GCCCAGCTGG TTTTCCAGT TTTTGCTGTT
8501 TGGACGGGGT GGCTGAGTAT GTTCTTCCAG GTCAATTGTCC TGTGCTGCCT
8551 TGCCTCCCTG AGCCTCTGTT TCTCCTGTTA AATGTTGATG ATTCCCTGCA
8601 TCCAGGCCTG GTTTAGAGGT GTGGTCTTT TGGCAGTGAG TATTGCTT
8651 AATTCAATGGC AATGAATTCA ATCCCCAGGG GCTGAGAGAG CCAGTCGTGG
8701 GGGACAGTAA GGGAGGTTT TACTCTTCA CCTGCTCCCTG ACCCTGACTC
8751 CTCCCTACCC CCTCCTACAT TTCCAGGGCT GAGGTAGGGA GGATAGTTGT
8801 GGGGGTATGA CTCCTCTGTC CTTTGTCCCC AGCTCTGGGA GGGTCTCACG
8851 GAACTAGTGA CGGCGACAGG CAACTATGGC AACTACCGGC GTCGGCTGGC
8901 AGCTGTGTT GGCTTCCGCT TCCCGATCCT GGGTGTGCAC CTCAAGGACC
8951 TGGTGGCCCT CGAGCTGGCA CTGCGTGAAT GGCTGGACCC AGCCCGGACC
9001 CGGCTCAACG GGGCAAGAT GAAGCAGCTC TTTAGCATCC TGGAGGGAGCT
9051 GGCCATGGTG ACCAGCCTGC GGCCACCAAGT ACAGGCCAAC CCCGACCTGC
9101 TGAGCCTGCT CACGGTGTAGG ACCAGGGGGC AGGGAGGTGG GGAGCTGGG
9151 ACCAGGGTT GACAGTTCC CCAGGTCTG GCTGTGGCG TGGCCTGGGG
9201 CTCTGGGTT TGGCCAAGAA ACTGAGATCT AGCGTGGGCT CTGGGGTT
9251 GAGTGGATGC TGAGAAGGGG TCCAGGCTCT GTGTTGGGCT GTGGACTGAG
9301 GTCTGATCTC CAGGCTGGTA TGTGGACTGT GGGCAGTTG AACTGGGCT
9351 GGGTCCCGGG TTGAGTTCTG GCAATGGGCT GTGTTCTAGG GCTGGGCAA
9401 GCTCTGCATT CTGTGGCAG GGGTGGTTTC TAAGCATGGC CCTGGCTCG

FIGURE 3, page 3 of 12

9451 GAGTGAAGTT CTGGGCTTGG CTTTACACTT GGTCTTGGGG TCTAGGGTGG
9501 GAGTTGGGTT CTGGTTAGA TCCAGACAAG GTTCTAGACA TTGGGCTGGG
9551 GCTTAAGTGT TAAGGTTGG AGTGGATTCT TAGCTGCTTC TGGGCTCTGG
9601 AGGGGATCAG GGTTGAAATC AGAGCTTCTG GCTGGGTTCC GACCTGGCTT
9651 CTTCCCTGAC ATCTTGCAA TATGTTGTGT TCAAGGTTTGGGGCATGCT
9701 GTGGTTTGTAT CTGTGCGCTG GGATGACATG GGGGTTGCTG TGCTGTGTC
9751 TAAGCCAGGC TTTGTCTGA GTCTAGCTTC TGACCCGAGC TCTGGCTGAG
9801 CTGTGGCTC TAGGTGACCC TTTGGGCCCTG GGCTCTGTGG CGTGGGCCAG
9851 GGGCCAGTGG GGGTGATCAG ATCTGTGTGT CCCAGGTGTC TCTGGATCAG
9901 TATCAGACGG AGGATGAGCT GTACCAGCTG TCCCTGCAGC GGGAGCCGCG
9951 CTCCAAGTCC TCGGTGAGGG GGTACTCCCT CCTCTCACT CTGCCCTTCC
10001 CTCCAGGAA TCCCAGGATG TGAGGATGGG AAGAGCTTTC AGCAGCCACC
10051 TCACCCATCC ATCTTGAGG ACAGAGGCAT CCTGGGGTGA GGGCAGTAGT
10101 GTTGGGCAGA CTTCCCTCTC CCAGGGATTCC CCTCTCTGT TCCCCGGGGC
10151 TCTGGGCTCC CCTCTGCCTC GGCCCTAGCT CAGGCCGAC CATTTCATA
10201 GCCAACCCAGC CCCACGAGTT GCACCCCCACC ACCCCGGGCC CCGGTACTGG
10251 AGGAGTGGAC CTCGGCTGCC AAACCCAAGC TGAGTCAGGC CCTCGTGGTG
10301 GAGCACATCG AGAAGATGGT GGAGGTGAGC TCCGTGCGAG CCTGAGCAGT
10351 GTGTGGGAG AGGCCAGTTT GCGGGAGCAC TGCCCCTGGA GCCAGCACCA
10401 GTGTCTGTGTT CAAGACCCAG CACTCAGCCC CTAGGAGTCAG CAGGGCCTGG
10451 CAGGCCAGCT GCACGGGGCT GAAGTGCCCC TGGGTAGGGT GGGGGTGGAG
10501 GTATGGAACG GGGGTGGTGT CAGAGACCTC TCTGAGACAC ACCTCATCAA
10551 ATGGACTGGG AACGTGGGAA GGGACAGGAC CTGATGTCCC CTTTACTCTC
10601 CCCTCTTCTG GCTCTGCGTG TCCCTCTGCG TGCCCCAGTC TGTGTTCCGG
10651 AACTTGACG TCGATGGGA TGGCCACATC TCACAGGAAG AATTCCAGAT
10701 CATCCGTGGG AACTTCCCTT ACCTCAGCGC CTTTGGGGAC CTCGACCAGA
10751 ACCAGTGAGG AGGGCTGGGG ACCTGGGGGA GAGGGAAAGGC AACTCAGCCC
10801 ACTTCTGCCT GGGCTTCAGT TTCTTGTTGTG CAAGATGAGG TCACTGAGCC
10851 AGATGATCTT GGCTGGGAA GCTGCCAGTG TGGGAAAGGG CACTTGCTTT
10901 TGTGGGAGG AGAGGCTGCC AGCTGTGGAG GCGCAGTGGT ATCTCACAAA
10951 TTCAGACAGA TGGGGGGCTC CACCTGAGTC TTGCAAAGAC TGTGACCTGG
11001 GGACTGTGGC TACAAAAGTG CTGTTTATT TGTGAGCTC ACAGCTGTCA
11051 AGAAAGTGTGG GCAACTTGAG CTCCCTGGATA GTCTGTTCTA ATGAATAGAT
11101 AAGAAAGTTT TGTAATTAGC AGTACCCAGT TGTTTATCAA CAGTTCATAT
11151 GCTGACAATT TGGAAAACA GCTGGTCTC TGAAGTAGGT TAAACATGCC
11201 CCCCTGAAGCC AGATTCTCATGC CCTATTTTTG CTGAGCAGAA AAAACTCCAT
11251 TCAAAATTAA AAGTCCATCT CAGGTCGATT TATTTTTAA TGTTACCTGT
11301 ATTCAAAAAA TCTGTTGTTT TTTATTCTCA CATTACAAAA ATCCACGGTA
11351 AAATAAAATC TAGGTGGTAA AATAAAATTAA TAGTGAACAA AATGTTAAA
11401 GTAAGAAGTG AGAGGCCAGG TGCGGTGCCT CACGCCCTGTA ATCCTAGCAC
11451 TTTGGGAGAC TGAGTTGGCA GGATCAATTG AGGCCAGGAG TTTGAGCCC
11501 GCCTGGGCAA CAGAGTAAGA CCCTGTCTCT ACAAAATTAA TTATTATTAT
11551 TTTGAGACA GAGTCTCACT CTGTTGCCA GGCTGGAGTG CAGTGGTACA
11601 ATCTCGGCTC GCTGCAACCT CCACCTCTG GGTCAAGTG ATTCTCCTGC
11651 TTCAGCTTCC TGAGTAGCTG GGATTACAGG CATGCATCAC CGTGCCTGGC
11701 TAATTTTTGT ATTTTTAGCA GAGATGGGT TTTACCATGT TGGCCAGGCT
11751 GGTCTCAAAC TCTTGACCTC AAGTGTATCTA CCTGCCCTGG CCCCCCAAAG
11801 TGCTAGGATT ACAGGCATGA GCTACTGCTC CTAGCCTAAA AAAATTTT
11851 TTGGGCATGG GTGGCACGTG CCTGTAGTCC CAGCTACTCA GGAGGCTGAG
11901 GCAGGAGGAA CCCTTGAGCC CAGGAGGTTG AGACTGCAGT GAGCTGTCA
11951 CACACCACTG CACTTCAGCC TGGGTGACTG CGCGAGATCA CCCCCATCAA
12001 AAAAAAAAAA AAAAGAAAAA AAAAGGAAGA AATGAAAGTC CCCCTTTCC
12051 TTTTCCACTG GTAGAAGITG CCATGATTAA GCACTGTTAA CAATATTAAAG
12101 CTTGGCAGTA TGTGGATTCT TCCAGTCTTC TTTTCCCAGG CAGGTGCACA
12151 TTGATAGAGA TTTTGTGTTGT TTGGTGTCTG TTTCATGGAC AAACAGGATT
12201 AGAGCATAAA TCTAGTTCTG CTTGTGGCTT TTATCATAGC TGCTTTATT
12251 CTTCTCCCAAG ATTTTAGGCA GAGGTAGTTG AGTCCCATGT TTTCTCCCTG
12301 GTTGGTGGGG TGGATTTTA TCTAGACACC CTTTCAGTG AGAATGACCC
12351 TTTGAGACGA TGGAGGCCCTC AGCTTCATGC AGCAGGGCTCA GCCTTAACCC
12401 TCCACCTCCT GCAGGCCCA AGCTGTGTGT GTGTGTGTGT GTGTGTGT
12451 GTGTGTGTGT GTGTGTGTGT GTTGGTAAGG GGAAGGCCAGC ACCGGCTCCC
12501 TCAAAACCT AGCACCTGGT TCGGCAGGAG GGAGACCAGC ACCGGCTCCC
12551 CAGGACCAGG CCCAGCTCAC CACTTCATTG TAAAGCTCCC TCTTTGTTTC

FIGURE 3, page 4 of 12

12601 TGGAACCTGG GTGTTTCCAT TTCTTTCTTA CAAAATTATC TATGCATTAA
 12651 CAGCAATTGT TGATATATCT TTAGGCAGCA TCTAGGTAATC TGTAGTGGGT
 12701 TCTCTTTTT CTTTTTCTT TTTTTTAATC ACCCTCTCTT TTTTTTGAGA
 12751 CAGAGTCTCA CTCTGTCGCT CAGGCTGGAG TGCAATAGCG CGATCTGGC
 12801 TCACTGCAAC CTCTGCCCTCC CAGGTTCAAG TAATTCTCAT GCCTCAGCCT
 12851 CCCAAGTAGC TGAGATTACA GGCACTGGCC ACCAGACCCG GCTAATT
 12901 TTTCTTTTT CTTTTTTTG AGACGGAGTT TCGCTCTTGG TTGCCCGAGG
 12951 TGGAGTACAG TGGTGTGATC TCGGCTCACT GCAACCTCCG CCTCCCCGGT
 13001 TCAAGTGATT CTCTGTCCTC AGCCTCCCAG GTAGCTGGGA TTACAGGCGC
 13051 GCGCCACCAT GCCTGGCTAA TTTGTATT TTTTTTTT GAGACAGAGT
 13101 CTCACTCTGT CACCCAGACT GGAGTGCAGGT GGCGCGATCT CGGCTCACTG
 13151 CAAGGCTCTGC TTCCGGTT CATGCCATTTC CCCTGCCCTCA GCCTCCGGAG
 13201 TAGCTGGGAC TACAACCAC CACCACCGTG CCGGGCTAAT TTTTGTTATT
 13251 TTTAGTAGAG ACGGGGTT ACCGTGGTCT CGACCTCCAG ACCTCGTGAT
 13301 CCACTAGCCCTCA AAGTGCCTGGG ATTACAGGCG TGAGCCACT
 13351 CACCCAGCCT AATTGTAT TTTTAGTAGA GATGGGGTTT CACCATGTTG
 13401 CGCAGGCTGG TATTGAACTT CTGACCTCAG GTGATCCGCC CGCCTCGGCC
 13451 TCCCGAAGTT CTGGGATTAT AGGCCTGAGC CACCGCACCT GGCTTAATT
 13501 TTGTATT TTTAGTAGAG GAGTTTACCT TTGTTGGCCA GGCTGGTCTT
 13551 GAACTCCTGA CCTCACCTCA GGTGATCTGC CCACCTCGGC CTCCCAAAGT
 13601 GCTGGGATTA CAGGCATGAG CCACTGTGCA CCCGGCTAA AAATCACCCT
 13651 CTTGACAGAA CTTCACGCC TGCTTTTGT TTTTTTCAT CTTTGTGCTT
 13701 GTTTTCCACT TAACCCCTGA TCACAGACAT CTTCCATGT GGATTCACTG
 13751 AGAACTACCT CATTGTTAG AACAGCTGCA GAGTATTCCA CTGTGCGGTT
 13801 AGTCATCAT TTCCCTAACC ATCCTCTGC TGATGGACAG TTAGACTGTT
 13851 CCAGTTTTC AGTATGATTC TATGCCAGGC TGCCATGAAC GTCCTTTAC
 13901 TGATCCACTC AGGCCAGTAT TTCTGTAGGA GAAATTCTA GAAGTGGGAT
 13951 AATTGGATCA AAAGATATGC ACATTCTAAA TTAGGAGAGA GACTGCCAA
 14001 CTGACCTCAG ACAAGGTTG ACCAGTTGC ACCCCCATCA GCAGCGTACA
 14051 AGTGCCTGCT TCCCAACTTC CTCGCCAACA GGGATGCTAT AAAAGCTTC
 14101 ACAATTTCGC CAGTCATGG CAATGGTAT TTCTGGTTA AATTGCAATT
 14151 TCTTAATAC TAAGTGGGGG TAGGGTATCT TTTCATATGT TTATTGGCCA
 14201 TTTATTCTT CTGTCATTG CCTGTTCTGA TTCTTGTCC ATTATTCTAC
 14251 TGGGTTGTT GGTCTTTTC TCATTGATT TTAGAATCTC TGTTAATGGA
 14301 TATTAACCCT TTGCTGTTGA ATGTGTTGC AAATATTTTC TCCCTGTCTG
 14351 TCATTTATGT GTCTTTTCC ATATAAATT AAAAATTTT GGTGGCTCA
 14401 ATAGGTCAGT CTTCCCTTC CGGGCTCTG GGATTTGTGT TCGGGTAGA
 14451 AAGGCCCTCA GCCCCTCAAG ATTATAAAAT TATAAAACCT TTCTTTTTT
 14501 TTTTTTTTCTG CTGAGACAGG GTGTCTGCC ATGTCACCCA GGCTGGAGTG
 14551 CAGTGGCATG ATCTGGCTC GCTGCAACCT CCACCTCCCA GGTCAAGTG
 14601 ATTCTCGTGC CTTAGCCTCC CGAGTAGCTG GGATTATAGG TGCCTGCCAC
 14651 TATGCCCTGGC TAATTTTTG TATTTTTAGT AGAGACGGGG CTTTGCCATG
 14701 TTGGCCAGGC TTGCTTGTAA CTCCCTGACCT CGTGATCCAC CCGCTTGCG
 14751 CTCCCAAAGT GCTGGACTA CAGGCGTAAG CCACGTGCT CGGCCCTATA
 14801 TTTTTTCAG ATAGCAGT ATCCTAATGC TCCCTTGATT TGATGGACCA
 14851 CCTGGATCAC ACATTATGAG CCCCCCTCATA AGCAGGTGGG AGTCTCAAGC
 14901 GAGGGCCAGT CCCGATGGGA ATAGCACTTG GTGGCTGAGG ACCCTCCTAT
 14951 CTGTGCAGAC ACTGTTGAA AACTTCACAT GCATCATCTA ATTTAGTCCCT
 15001 CACCAAAATC CTATGAAATG TAGGAATGAT CATTACACCC ATTTATAGAT
 15051 AAGGAAACGG AGGGACAGGG AGATTACTCC GCTACAGGTC AAGAGGCAGG
 15101 GAAAGTAGAGC TGCGATTGAA ACTGAGGTCT GTGTCTAGAA CACGTGCTCA
 15151 TTCTTCCCT AAAATGTATT CATAGGTGAA AAAGGGCTTC TGCAGGAAAGC
 15201 CCTGGGTTAT GTGGAAACC CTGGATTAC AGCTGTCTT CCAGCAGGAT
 15251 GATGCAGGAG AGAGAGGGAT GCGATTCTC CCAATCTCTC CTGGTCCAG
 15301 AACTCATTAG AGAGTTCTCC CTGCTGAGGG CTCCCGACTG GTGTTGCACA
 15351 CAGTACACTT CGGGAGCCCG AGGCTGATGG TTCCATGGAA AGTACACAGT
 15401 CATTTCAGTT TGACACCAA GTGTGAAGTG GGCAAGGACAG GCCACTGTT
 15451 TGAGAAGGAA CCCAGGGAAA GGGACTGGCC CAAGACCACA CACTGGTAG
 15501 CGGCACTTCC CACATCTGCC TGACCCCTAG TCCAGTGCCG CCTTTCTT
 15551 ACTCTGCAAC AGGAGTCCAA AATCAGGAGT TCCATGAGGA CACTGGGAAC
 15601 AGTGGGATGG GTTACGGCCAG CGGTGGATGG TTCTGGGGAG GGCCCGAGCT
 15651 GAAGCGCCCC CGCAACTCCC CACAGGGATG GCTGCATCAG CAGGGAGGAG
 15701 ATGGTTTCCT ATTTCCTGCG CTCCAGCTCT GTGTTGGGGG GGCGCATGGG

FIGURE 3, page 5 of 12

15751 CTTCGTACAC AACTTCCAGG AGAGCAACTC CTTGCGCCGC GTCGCCTGCC
 15801 GCCACTGCAA AGCCCTGGTG AGAGTCCTT TCCCGGCTCA CGGCCAAAGC
 15851 CACGCCCTC CAGCCCGGC CCCGCCCTCC CTCTCTGCC CGCCTCTGCC
 15901 AGAGCCCTTC TCAAGCCAGG AAAACCTGGT AATTCTATTT GCCTCTCC
 15951 CTGTGGTTCT GCCCCGGGCT CTGAGGGGG CTCTAAAGCC CTAGTCTCAC
 16001 CCTCAAGAAG GAAGAAGTAG AGTCATCACC TCTAAATCCC TCCTCCCACC
 16051 ACAGCCCTC CTCTATTGCA GATCCTGGGC ATCTACAAGC AGGGCCTCAA
 16101 ATGCCGAGGT GAGATGGAAG GACTGGAAGG GCTGCTGGC AGTGTTTTT
 16151 TTGTTGTTT GTTGTGTTG GAGAGTTACT ATTTGGTGG GGCAATTGCC
 16201 AAGGAGTGAATGAA ATCACCTTAAATCAGAGGCG CATGGCCGGG CATGGTGGCT
 16251 CAAGGCTGTA ATCCCAGCAC GCCTGACCAA CATAAGCGCAA CCCCACCTCT
 16301 AGGTCAAGGAG TTCAAGACCA GCCTGACCAA CATAAGCGCAA CCCCACCTCT
 16351 ACTAAAAATA CAAAAAAGTAG CTGGGCGTGG TGGCACCCAC CTGTAATCCC
 16401 AGCTACTTGG GAGGCTGAGG CATGAGAATC GCTGAAACCT GGGAGGCGGG
 16451 GTTTGCAGTG AGCCGAGATC ACGCCACTGC ACTCCAGCCT GGGCAACAGA
 16501 GAGGGCTCTG TCTCAAAAAA AAAAAACAAAC AAAAAAACCC CCAAAACCAA
 16551 AACCCCACAA ATCAGAGGC TCAAGATGAC TGATGTGAAG GGAGTGGCGT
 16601 TTAAGAGGCC ATTATTTTG ATGACGCAGC TGCCCAGGAA CAGAGAACAT
 16651 GGGAGAAGGC ATAGACTGAC ATTAGGAGG AGGAGAACAC TTTGGAAGGA
 16701 GACTCTTATT TTGGTGGGGC AGCTGCTAG GAACAAAGGT TCCTGGTAGG
 16751 GGGCGCAAG CCTGCGGGAT GGGATGGAGG GTATTCTGAC CAATGTCC
 16801 GGCTGGCTCT CCATTTGCTC TCCCCCAGCC TGTGGAGTGA ACTGCCACAA
 16851 GCAGTGCAAG GATCGCCTGT CAGTTGAGTG TCGGCGCAGG GCCCAGAGTG
 16901 TGAGCCTGGA GGGGTCTGCA CCCTCACCC CACCCATGCA CAGCCACCAT
 16951 CACCGCGCCT TCAGCTTCTC TCTGCCCGC CCTGGCAGGC GAGGCTCCAG
 17001 GCCTCCAGGT AAGAGGGAGT CATTCTGTAC TGGCTGTGG AGGGAAGGAT
 17051 GCAGGGCTAC TGGGGCAAAG AACGCAGGAT GGAAGCCATT CCAAAGTGCA
 17101 TAATTCTCTT TTGTGGTGG GATAATAAAG AAGGGACAGG CCGGGCGCGG
 17151 TGGCTCACGC CTGTAATCCC AGCACTTGG GAGGCGAGG CGGGCGGATC
 17201 ACGAGGTCAG GAGATCGAGA CCATCTGGC TAACACGGTG AAACCCCATC
 17251 TTTACTAAAA ATACAAAAA AAAAAATTAG CCAGGCGTGG TGGCGGACGC
 17301 CTGTAGTCCC AGCTACTTGG GAGGCTGAGG CAGGAGAATG GCATGAACCC
 17351 GGGAGGCGGG CTTGCACTG AGCCGAGATC GCGCCACTGC ACTCCAGCCT
 17401 GGGCGATAGA GCAAGACTCC GTCTCAGAAA AAAAAAAAT AAAAAATAAA
 17451 GAAGGGACAG GTAAGGGTGC CAGAAAGTGG CCAGGAAGGC CTGGACCTTC
 17501 TGAGGCTGAG GAGAGAGACC CTAATTATA AAGAGGTATA AAAGTGAAG
 17551 AGGCTTCAAG ATTCCAGTTA CAGTCTTATT TTGTTGGAGG GTTTAACAAA
 17601 GGATTGGAGA AGGTGTTATA TGAGCCATTG GCTTGCCTTT CCCTTCTGG
 17651 CTGCTCTGGA GGCTCTTCTG GGGAAAGTCC CTTGCCCTGA TAATGTCTC
 17701 GCAGCTCTCT TGGGGTATTG GATGGTTTA GGTCACTTTC CTGAATGACA
 17751 ACTGGCCAAA TGATTATTT GCTGAGAACAA GTCCGAACAA CTATGTTAAA
 17801 CTGGGGTCTA AGGTAGTTGA TCACAACACTG TTGGGTTGGC ATAAGTCTC
 17851 AAAAAACAGA GGCAGGCACA GGGCATAACAT CCTCAAAAT AGAAAAGATA
 17901 AATCCATTG CATTGAGCT TCCAGAAGTG CTGGGGTCTA AAATGTGAAA
 17951 TACACACAA ATTGACATT AAGCAAATG CGCTGACAAA TCTGTGGCTG
 18001 AAAAAGCTGT GGCAAAACAA AAACATAGAA AAAGAGCCTC AAAAATTGGG
 18051 CTGAGGCCAG GCATGGTGGC TCACGCCCTGT AATCCTAGCA CTTTGGGAAG
 18101 CCAAGGTGGG TGGATCACCC GAGGTCAAGGA GTTGGAGACC AGACTGGCCA
 18151 ACGTGGCAAA ACATCATCTC TACAATACAA AAATACAAA ATTAGCTGGG
 18201 CGTGGTGGCA GGCGCCTGTA ATCCCAGCTA CTGGGAGGC TGAGGCACGA
 18251 GAATCGCTTG AACCTGGGAG GTGGAGGTTG CAGAGAGCCG AGATTGCC
 18301 ATTGCACTCC AGCCTGGCG ACAGAGAGAG ACTCTGTCTC AAAAAAAA
 18351 AAAAAAAA AAATTGGGCT GTGAGGTCA GCAAGGAAATT GATTTGGGT
 18401 GGGTGGGTCT GCTTCTGGGA TGATGTGGAT GCCTCCCGTG GAGAGGGAA
 18451 GGGTTGATGA AGTCCCAGGG ACCTGGAAGT GTGTTCTGCA GCAATCCCC
 18501 TCCCAGCAGA GATCCGTGAG GAGGAGGTAC AGACGGTGGA GGATGGGTG
 18551 TTTGACATCC ACTTGTAAATA GATGGTGAGT CCTCCCACAG CTGGCACAG
 18601 AGCTCCCCAC TGAGGGCTGG GGGGGAGCTG GGGAGTATCA GGGAAATGGG
 18651 TGCTTTATCC AAATGGCTCC AAGCCAGGTG GGCTACTACC TTGTTGTAG
 18701 GGGGGTGTCT TCCTCACAAC CTGTTTTCT CTTCCCAGCT GTGGTTGGAT
 18751 CAAGGACTCA TTCCCTGCCTT GGAGAAAATA CTCAACCCAG AGCAGGGAGC
 18801 CTGGGGGTGT CGGGGCAGGA GGCTGGGGAT GGGGGTGGGA TATGAGGGTG
 18851 GCATGCAGCT GAGGGCAGGG CCAGGGCTGG TGTCCCTAAG GTTGTACAGA

FIGURE 3, page 6 of 12

18901 CTCTTGTGAA TATTTGTATT TTCCAGATGG AATAAAAAGG CCCGTGTAAAT
 18951 TAACCTTCAC CATCAGCGCC TAGAACATCCG GGGGGTAGGG GGATGGTATA
 19001 CTTTACAGGA TGACAATCTT GGGAGCTAGA ACITTGATGC CAGAGAAACT
 19051 TGGGAGGTCT GGAATCTCAT GTGTCTGGAG TCTTGGGGAA GAGAATCTTA
 19101 GAAGCAGAAA ACCTTGGAAC ATAAGAATCT TGGGGAGGGT CTAGGATCTT
 19151 GAGGAGACCA GATCCTTGGA CATCTAAAAC TTGAAACTAG TAGGTCTGCA
 19201 CCCGAGAATT GCAGGGCCAG TCATGCATAC CCAAAGGCCCTT CAGCCCCATGG
 19251 CCGAAATTCC CTTGCTGGAC AGGGGGCCCTT TCAGCCCCCTG CTTGGACGCT
 19301 TCCAGTAACA GGGCCCTCAC TGCAGGAATC GTGGGAGGGA GAGGGGCAGC
 19351 ACAGAGTTGC TGGCTGTCGG GGAAGGGAGG GAGGGCCCTG GGCAGTCCGA
 19401 GGGCCTGCT GGGCTTGTGCG CTCAGGGTGG GGGCTGCCT CCTCCGCCTT
 19451 GCAGGCTCCT GGCCTGGTGC TGCTGCCAGC CGGAAGGACA GTGACTTCCA
 19501 GAGGAATATGC ATATTGATCC TGCTTCAGC CTCCGGTGGT GGCTTCTCCC
 19551 AACCCAGCTC TTCCCTCCTG AGCCTGCAGC ACGGAGGTTT TGGGGTCAAC
 19601 TGCTACCTAA AGAAGGCTAA GGCCACTTCT GAGGCTGGTC TGGGAGTTA
 19651 CTAAAGGTTTC TGAAGCTGGG CGGGGCTGCC CCTGGGATCA GGAGACTCCA
 19701 GACAGCAGTC CTGACAATGG GAACTACCTC CTCAGTCCCC CAAACTGGGA
 19751 GGTGTCCCAC AGCAGCTGTA GGATTGTCTT AGGGGTGGAG ACCTGAGCAC
 19801 CTTCACTCC AAAGCACAGT ATCTGTGGGCT CGGCAGTGG CCTCAGTTCC
 19851 CCCATGAGTG CCCCAGGCTCC CCACCCAGG GTTCCCCAC ATCACATCCA
 19901 TCCCTGCTTT GAGACCCAC TCCCCCTGGC CTGTTCTTA TTTTGGGTC
 19951 CTCCCTCTC TTTCCTGGTC ATATCTCTCC TGCAAGGCCTA CCCTGTGTTG
 20001 GGCCCCCAG CCCTGTCTCT GCATCGGGTG CCCCCCTGCC CCTCCTCTG
 20051 TCCTCAGGCC CCTCCGCCCT TCCCCCTCTT GAGGCTGTAA TATCCGTTTC
 20101 ACGATTGGGG GGCTGAGTTG CTATAACAAAC AGACGGCGAT TGTGTTGTGA
 20151 AGAGCAGCTC GCTCCTGTGC CGCCTGCCCT CTGTGCTGCC TCCATCCCTG
 20201 CAGCCCAGTC GGTTCCCTCTT GGCTCCTCTC GTCACTACCC TCCAGTTCCA
 20251 GTCTGGCCTC TTCCCTGGTGT GTGTGTGTGT GTGTGTGTAT
 20301 GCATGCATGC ATATGTGTGT CCAGGCTCTGC CTGCCCGGGG TGTGACAAGT
 20351 AGCGGTCTTC ATGGTTGCAT GTGTCTGAAT TTGGTGTCTG AGCTTACACAT
 20401 TGTATGCGCC TGTGTGCATG TGTGTGCATG GACATGCATG CTGTATCTGC
 20451 TGTGTTCCC CTCCCCCATG TGTCCCCACT GCCCTTGCA CATGGGAGAA
 20501 GGGCATGTGC TCAGCATATC ACTCAACTGT CCACATTGGG TGGGTACCTG
 20551 TGTGTTGGTGT GTGTGTGTGG GGGGTGTGTC TTGAAGTGGC AGGTCCAAA
 20601 TGCTTAGGCA ATCTGAACCT TGGACCTTPGC AGAGAGGAGA GATGTCCCTG
 20651 TAGTGGGAG GGACAGGGAG ATGCAGCAGC TGCCCGGTGA CCTTTCTGC
 20701 CCTTGATGGG CAAAGCTGGG GGTAGGGAAA GGAGACAAGT GCTCATACTT
 20751 ACCCTCCCTCC CTGCCAGGC TCCTCTGTAA GGGTCTGAGT CTGTCTCTGT
 20801 GAGCCATTGC ATCTGTCTGT CTATGCCCTG ATGCCTGGAT GGACAAGGGG
 20851 TGTGTGTGTG TGTGTGTGTG TGTGTGTGTG AGTGTGAGGC TGCAGGAAGA
 20901 GGAACAGTGGG GGGATGGGCA GGAAAGTGGG CTGTGGGGTC AGGGAGGCAGA
 20951 T (SEQ ID NO:3)

FEATURES:

Start: 3000
 Exon: 3000-3072
 Intron: 3073-3753
 Exon: 3754-3855
 Intron: 3856-4363
 Exon: 4364-4427
 Intron: 4428-4786
 Exon: 4787-4918
 Intron: 4919-5702
 Exon: 5703-5853
 Intron: 5854-6056
 Exon: 6057-6230
 Intron: 6231-6389
 Exon: 6390-6506
 Intron: 6507-8832
 Exon: 8833-9114
 Intron: 9115-9885
 Exon: 9886-9963

Intron: 9964-10201
 Exon: 10202-10324
 Intron: 10325-10638
 Exon: 10639-10754
 Intron: 10755-15675
 Exon: 15676-15817
 Intron: 15818-16071
 Exon: 16072-16108
 Intron: 16109-16828
 Exon: 16829-17008
 Intron: 17009-18491
 Exon: 18492-18565
 Stop: 18566

CHROMOSOME MAP POSITION:

Chromosome 11

ALLELIC VARIANTS (SNPs):

DNA

| Position | Major | Minor | Domain |
|----------|-------|-------|----------------|
| 5539 | C | G | Intron |
| 5658 | T | G | Intron |
| 5861 | C | T | Intron |
| 6023 | A | G | Intron |
| 6799 | C | T | Intron |
| 9579 | C | A | Intron |
| 9842 | T | C | Intron |
| 10159 | T | C | Intron |
| 12025 | A | - G | Intron |
| 14723 | T | C | Intron |
| 14996 | G | A | Intron |
| 16153 | T | G | Intron |
| 16181 | G | A | Intron |
| 16756 | A | G | Intron |
| 18059 | A | G | Intron |
| 18364 | A | - T | Intron |
| 18861 | G | A | Beyond ORF(3') |
| 20443 | G | A | Beyond ORF(3') |
| 20881 | A | T | Beyond ORF(3') |

Context:

DNA

Position

| | |
|------|--|
| 5539 | AGACTCAGTCAGGCATGAAGTCCTCGTGGGCTCTGAGGGTTGGGGCTTTCCGGGTA GAATTTGTCGTTCCCACCTCTGTTTCCATGGCACTTTGTACAGACTCCTGTACAAAGAC CTCTGTACATGTGTACGCTGTTTGATCATGTGTTCTGTGTCTGTCTCCCTCAGTA GACTGTGAGCTCTCGAGGGCAGGAACCGTGTCTACTCATCTGTATTCCCAGCGCCT AGCACAGTCGCTGGCACAGAGTACGTTGTTCATAAATGTGTGTTGAGTGCATGACGGGT [C, G] GGGGGAGATGAGGAGGAGTTGCTGGACTGGAACATTCTGTGCCTAGGACAGTGCTCGC ATTATGTAGGTTCTCAGTAAGCGTGAATGGTGTCTGTGTGAGTGGGGGCCACGAGGC ATGCGCATGTCCAGCAAAGGGCTCACTACCCCTGCCCCCCCAGCCCTACCTACAAGTGG AGCGGCAGGTGACTCAGCGAACCTGTGGGACAGAAAAGCGCAAGATGTCCCTGTTGT TTGACCACCTGGAGGCCATGGAGCTGGCGGAGCATCTCACCTACTTGGAGTATCGCTCCT |
| 5658 | CCTCTGTACATGTGTACGCTGTTTGATCATGTGTTCTGTGTCTGTCTCCCTCAGT AGACTGTGAGCTCCTCGAGGGCAGGAACCGTGTCTACTCATCTGTATTCCCAGCGCC TAGCACAGTGCCTGGCACAGAGTACGTTGTTCATAAATGTGTGTTGAGTGCATGACGGG TGGGGGGAGATGAGGAGGAGTTGCTGGGACTGGAACATTGTGCCTAGGACAGTGCTC GCATTATGTAGGTTCTCAGTAAGCGTGAATGGTGTCTGTGAGTGGGGGCCACGAG |

5' → 3' 3' ← 5'

| | |
|------|---|
| | [T, G] |
| | CATGCGCATGCCAGCAAAGGGCTCACTACCCCTGCCCCCCCAGCCCTACCTACAAGTGG AAGCGGCAGGTGACTCAGCGAACCCGTGGGACAGAAAAGCGAAGATGTCCCTGTTG TTTGACCACCTGGAGCCATGGAGCTGGCGGAGCATCTCACCTACTTGGAGTATCGCTCC TTCTGCAAGATCCTGGTGCAGGGCTGGGGGTCAAGGGTCCAATGTGGCTGGA AGAGAGTTCTAGGAGGGGAGGGTCCCTGGCGTAGGCTGGTCAAGGGTGCATCAGGGG |
| 5861 | CTGGGACTGGGAAACATTGCGCTAGGACAGTGCCTCGCATTATGTAGGTTCTCAGTAAG CGTGAATGGTGTGTCGAGTGGGGGCCACGAGGCATGCGCATGTCCAGCAAAGGG CTCACTACCCCTGCCCCCCCAGCCCTACCTACAAGTGGAAAGCGGCAGGTGACTCAGCGGA ACCTGTCGGACAGAAAAGCGAAGATGTCCCTGTTGACCACCTGGAGGCCATGG AGCTGGCGGAGCATCTCACCTACTTGGAGTATCGCTCCTCTGCAAGATCCTGGTGCAGG [C, T] CGAGGGCTGGGGGTCAAGGGTCAAATGTGGGCTGGAAGAGAGTTCTAGGAGGGCAGGG TCCCCTGGCGTAGGCTGGTCACAGGGTCATCAGGGGTTTCAGTGTAAACACTGAAGGTC AGCTGGAGGGTGGAGGTGGCTATCAGTGAGGGAGAGGCCAAGGGTGTGAGGCCAC TCCTCATGCCCTCAGTTCAAGGACTATCACAGTTGACTCATGGCTGACTGTGGAC AACCCGTCCTGGAGCGGTCATCTCCCTCTTCAACAGCGTCACTGGGTGCAGCTC |
| 6023 | GGCAGGTGACTCAGCGAACCCGTGGGACAGAAAAGCGAAGATGTCCCTGTTGTTG ACCACCTGGAGGCCATGGAGCTGGCGGAGCATCTCACCTACTTGGAGTATCGCTCCTCT GCAAGATCTGTCGGCCGAGGGCTGGGGGTCAAGGGTCAAATGTGGGCTGGAAGAG AGTTCTAGGAGGGCAGGGTCCCTGGCGTAGGCTGGTCAAGGGTGCATCAGGGGTTTC AGTGTAAACCACTGAAGGTCACTGGGAGGGTGGCTATCAGTGAGGGAGAGGCC [A, G] GCAAGGTGCTGAGGCCACTCCTCATGCCCAAGTTCAAGGACTATCACAGTTGACT CATGGCTGCACTGTGGACAACCCGCTGGAGCGGTCATCTCCCTCTCAACAGCGTC TCACAGGGTGAGCTCATGATCCTCAGCAAACCCACAGCCCGCAGCGGGCTGGTC ATCACACACTTGTCCACGTGGCGAGGTGCCTGCCCTCCCTCCGGTGTCTCCCAACC ACCCACATGCCAGTCAGGCAACCCCTCCCTCCCTAACCACTGCCTCTCTAGA |
| 6799 | CCATCAAGGTGCCGGACTGGGAGGGGCCGTGCTTCCAGGTCTGCTTCAGTGGT CCTCCCAGCAGCACTGGGGCTGGCACAGCTGCTCTCATTTGATAGATATGAAATGGA GGCTCAGAGGGTTAAGTGTCTTCTCAGTTGACAATGGCAACAGCAGAGTGGGGCT CACAGGTGTCAGGGACCAAGCTAGTACTTTTTTTTTAAAGACAGGGTCTC TCTCTGTTGTCAGACTGGAGTTCAAGTGGTGCAGTCACAAGCTCACTGCAGCCTTGAA [C, T] TCCTGAGCTCAATCGATCCTCCCACCTCAGCTCCTGAGTAGCTGGGACTACAGGTGTAC GCCACCATGCCATAATTGGTATTGTTAAATTGTTTTTTTTAGAGATGGGG TTTGCCATGGCCAGACTGGCTTGAACCTCTGGCTCAAGTGTGATCCGCTGCTTG GCCTCCCAAAGTGTGAGATTATGGCTTGAGCCATTGTGCTTGCACATTGAGTTCTT CTTTCTTCTCCTCATTTTATTATTTGAAGTATTGAGTAACATA |
| 9579 | CTGGTTGGGCTGTGGACTGAGGTCTGATCTCCAGGCTGGTATGTGGACTGTGGCAGTT TGAACCTGGCCTGGGCTCCGGGTTGAGTTCTGGCAATGGGCTGTTCTAGGGCTGGCC AAGCTCTGCAATTCTGTGGGAGGGGTGGTTCTAAGCATGGCCCTGGGCTGGAGTGAAG TTCTGGCTTGGCTTACACTTGGCTTGGGCTAGGGTGGGAGTTGGGTCTGGTTA GATCCAGACAAGGTTCTAGACATTGGCTGGGCTTAAGTGTAAAGTTGGAGTGGATT [C, A] TTAGCTGCTCTGGGCTCTGGAGGGATCAGGGTCAAATCAGAGCTCTGGCTGGGTT CGACCTGGCTCTTCCCTGACATCTGGCAATATGTTGTCAGGTTGGGCCATGC TGTGGTTGATCTGTGCGCTGGGATGACATGGGGTTGCTGTGCTGTTCAAGCCAGG CTTGTCTGAGCTAGCTCTGACCCGAGCTGGCTGAGCTGTGGCTCTAGGTGAC CTTGCCCTGGGCTCTGTGGCCGTTGGGAGGGCCAGGGGGTGAATCAGATCTGTGTG |
| 9842 | TGGGCTGGGCTTAAGTGTAAAGTTGGAGTGGATTCTAGCTGCTCTGGCTCTGGA GGGGATCAGGGTTGAAATCAGAGCTCTGGCTGGGTTCCACCTGGCTCTTCCCTGACA TCTTGGCAATATGTTGTTCAAGGTTGGGGCATGCTGTTGATCTGTGCGCTGG GATGACATGGGGTTGCTGTGCTGTTCAAGCCAGGTTGCTCTGAGTCTAGCTTCT GACCCGAGCTCTGGCTGAGCTGTGGCTCTAGGTGACCTTGGCCCTGGCTCTGTGGC [T, C] GTGGGCAGGGGCCAGTGGGGTGTACAGATCTGTGTCAGGCTCTGGATCAGTA TCAGACGGAGGATGAGCTGTACCACTGTCCTGCAAGCGGGAGCCGCGCTCCAAGTCCTC |

FIGURE 3, page 9 of 12

| | |
|-------|--|
| | GGTGAGGGGGTACTCCCTCCTCCACTCTGCCCTCCCTGAGAATCCCAGGATGTG AGGATGGGAAGAGCTTAGCAGCCACCTCACCCATCCATCTGTAGGACAGAGGCATCC TGGGGGTAGGGCAGTAGTGTGGCAGACTTCCTCTCCAGGGATTCCCCTCTGTTC |
| 10159 | GGGGGTGATCAGATCTGTGTCCCAGGTGTCAGTATCAGACGGAGGTGAG CTGTACCAGCTGCCCTGCAGCGGGAGCCGCGCTCCAAGTCCTCGGTAGGGGTACTCC CTCCCTCCACTCTGCCCTCCCTGAGAATCCCAGGATGTAGGATGGGAAGAGCTC TTAGCAGCCACCTCACCCATCCATCTGTAGGACAGAGGCATCTGGGGTAGGGCAGTA GTGTTGGCAGACTTCCTCTCCAGGGATTCCCCTCTGTTCCTCCCCGGGCTCTGGGCT [T, C] CCCCCTGCTCTGCCCTAGCTCAGGCCCGACCACTTCATAGCCAACCCAGCCCCACGGAT TGCACCCACCACCCGGCCCCGGTACTGGAGGAGTGGACCTCGGCTGCCAACCCAAG CTGGATCAGGCCCTCGTGGTGGAGCACATCGAGAAGATGGTGGAGGTGAGCTCTGCCGA GCCCTGAGCAGTGTGGGGAGAGGCCAGTTGCCGGAGCACTGCCCTGGAAGCCAGCACG AGTGTCTGTTCAAGACCCAGCACTCAGCCCCCTAGGAGTCACAGGGCTGCCAGGCCAGC |
| 12025 | TGGGGTTTACCATGTTGCCAGGCTGGTCTAAACTCTGACCTCAAGTGTACCTG CCTTGGCCCCCAAAGTGTAGGATTACAGGCATGAGCTACTGCTCTAGCCTAAAAAAA TTTTTTTGGCATGGTGGCACGTGCCTGTAGTCCCAGCTACTCAGGAGGTGAGGCAG GAGGAACCTTGAGCCAGGAGGTGAGACTGCAGTGAGCTGTATCACACCACTGCACT TCAGCCTGGTGACTGCGAGATCACCCCATCAAAAAAAAAAGAAAAAAA [A, -, G] GAAGAAATGAAAGTCCCCTTTCCACTGGTAGAAGTGGCATGATTAAGCACT GTTAACAAATATTAAGCTGGCAGTATGTGGATTCTCCAGTCTTCTTCCCAGGAGGT GCACATTGATAGAGATTTGTTGTTGGTGTCTGTTCATGGACAAACAGGATTAGAGC ATAAAATCTAGTTCTGCTGTGGCTTTATCATAGCTGTTTCTTCTCCAGATT AGGCAGAGGTAGTTGAGTCCATGTTCTCCCTGGGTTGGTGGGATTTTATCTAG |
| 14723 | GGCTTCTGGATTGTGTTGGGTAGAAAGGCCCTCAGCCCCCTCAAGATTATAAATTA TAAACCTTTCTTTTTTTCTGAGACAGGGTGTCTGCCATGTCACCCAGG CTGGAGTGCAGTGGCATGATCTGGCTCGCTGCAACCTCCACCTCCAGGTTCAAGTGT TCTCGTGCCTTAGCCTCCGAGTAGCTGGATTAGGTGCCTGCCACTATGCCCTGGCTA ATTTTTGTATTTAGTAGAGACGGGCTTGCCATGTTGCCAGGCTGGTCTGAACT [T, C] CTGACCTCGTATCCACCGCCTGGCTCCAAAGTGTGGACTACAGGCGTAAGCCA CTGTGCTGGCCCTATATTTTCAGATAGCCAGTTATCCTAATGCTCCCTGATTGA TGGACCACTGGATCACACATTATGAGCCCCCTCATAGCAGGTGGAGTCTCAAGCGAG GGCCAGTCCGATGGAAATAGCATTGGTGGCTGAGGACCTCCTATGTGCAGACACT GTTGAAACTTCACATGCATCATCTAATTAGTCCTCACAAAATCTATGAAATGTAG |
| 14996 | CCATGTTGGCCAGGCTGGTCTGAACTCCTGACCTCGTGTATCCACCCGCTTGGCTCCC AAAGTGTGGACTACAGGCGTAAGCCACTGTGCTGGCCCTATAATTTCAGATAGC CAGTTATCCTAATGCTCCCTGATTTGATGGACCCACCTGGATCACACATTATGAGCCCCC TCATAAGCAGGTGGAGTCTCAAGCGAGGGCAGTCCCGATGGGAATAGCACTTGGTGGC TGAGGACCTCCTATCTGTGCAGACACTGTTGTAAAACTCACATGCATCATCTAATT [G, A] TCCTCACAAAATCTATGAAATGTAGGAATGATCATTACACCCATTATAGATAAGGAA ACGGAGGGACAGGGAGATTACTCCGCTACAGGTCAAGAGGCAGGGAAAGTAGAGCTGCGAT TTGAACTGAGGTCTGTGCTAGAACACGTGCTCATTCTTCCCTAAAATGTATTGATAGG TGAAAAAGGGCTTCTGGAAAGCCCTGGGTATGTGGGAAACCTGGATTACAGCTGT CTTCCAGCAGGATGATGCAGGAGAGAGAGGGATGCGATTCTCCAAATCTCTGGTC |
| 16153 | CGCCCCCTCCAGCCCCGGCCCCGCCCTCCCTCTGGCCCCGGCTCTGCCAGAGCCCTCTC AAGCCAGGAAAACCTGGTAACTCTATTGCTCTCCCTGTGGTTCTGCCGGGGCCCT GAGGCGGGCTCTAAAGCCCTAGTCTCACCCCTCAAGAAGGAAGTAGAGTCATCACCTC TAAATCCCTCCACCGGCCCTCCTCTATTGCAAGATCTGGCATCTACAAGCAG GCCCTCAAATGCCGAGGTGAGATGGAATGACTGGAAGGGCTGCTGGCAGTGT [T, G] TTTGTGTTGTTGGGAGAGTTACTATTGTTGGGGCAATTGCCAAGGAGTGAAGTA CTTAAATCAGAGGCGCATGGCGGGCATGGTGGCTCAAGCCTGTAATCCACCACTTT GGGAGGCCGAGGCCGCGCAGATCACCTGAGGTCAAGGAGTCAAGACCAGCCTGACCAACAT AGCGCAACCCCGCCTACTAAAAATACAAAAGTAGCTGGCGTGGGCACCCACCTG TAATCCCAGCTACTTGGGAGGCTGAGGCATGAGAATCGCTTGAACCTGGAGGCGGGTT |

FIGURE 3, page 10 of 12

bioRxiv preprint doi: https://doi.org/10.1101/2022.09.22.501362; this version posted September 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

| | |
|-------|---|
| 16181 | CTTCTGGCCCCGCCTGCCAGAGCCCTCTCAAGCCAGGAAACCTGGTAATTCTATTT GCCTCTCCTCCTGTGTTCTGCCCGGGCCCTGAGGCGGGCTCTAAAGCCCTAGTCAC CCTCAAGAAGGAAGAAGTAGAGTCATCACCTCTAAATCCCTCCACCACGGCCCTC CTCTATTGCAGATCCTGGCATCTACAAGCAGGGCCTCAAATGCCAGGTGAGATGGAAT GACTGGAAGGGCTGCTGGCAGTGTGTTGTTGTTGTTGGAGAGTTACT [G, A] TTTGTTGGGCAATTGCCAAGGAGTGAAGTACCTAAAATCAGAGGCCATGGCCGGC ATGGTGGCTCAAGCCTGTAATCCCAGCACTTGGGAGGCCAGATCACCTGA GGTCAGGAGTTCAAGACCAGCCTGACCAACATAGCGAACCCGCCCTACTAAAAATAC AAAAAGTAGCTGGCGTGGCACCCACCTGTAATCCAGCTACTTGGGAGGCTGAGGC ATGAGAACATCGCTTGAACCTGGGAGGCCGGGTTGCACTGAGCCGAGATCACGCCACTGCA |
| 16756 | CAGTGAGGCCGAGATCACGCCACTGCACCTCCAGCCTGGCAACAGAGAGGGCTCTGCTCA AAAAAAAAAAAACAACAAAAACCCCCAAAACCAAAACCCCACAAAATCAGAGGCTCAAG ATGACTGATGTGAAGGGAGTGGCTTTAAGAGGCCATTATTTGATGACGCAGCTGCC AGGAACAGAGAACATGGGAGAAGGCATAGACTGACAATTAGGAGGAGGAAACACTTGG AAGGAGACTCTTATTTGGTGGGCAGCTGCTCAGGAACAAAGGTCTGGTAGGGGGC [A, G] CAAGCCTGCGGGATGGGATGGAGGGTATTCTGACCAATGTCCTGGCTGGCTCTCCATT GCTCTCCCCCAGCCTGAGGACTGCCACAAGCAGTCAAGGATGCCCTGTCAGTT AGTGTGCGCGCAGGGCCAGAGTGTGAGCCTGGAGGGCTGCAACCTCACCCCTCACCA TGCACAGCCACCATCACCGCGCTTCAGCTCTCTGCCCCGCCCTGGCAGGCCAGGCT CCAGGCCTCCAGGTAAAGAGGGAGTCATTCTGACTGGCCTGTGGAGGGAAAGGATGCAGGG |
| 18059 | AATGATTATTTGCTGAGAACAGTCGAACAATATGTTAACTGGGTCTAAGGTAGTT GATCACAACCTGTTGGGTTGGCATAAGTCCTAAAAACAGAGGCCAGGGCAGGGCATA ATCCTAAAAATAGAAAAGATAAAATCCATTGCAATTGACCTTAAAGCAAACGCGCTGACAATCTG TAAAATGTGAAATACACACAAAATTGACATTAAAGCAAACGCGCTGACAATCTG TGAAAAAGCTGTGGCAAAACAAAACATAGAAAAGAGCCTAAACATTGGCTGAGGCC [A, G] GGCATGGTGGCTCACGCCGTAACTCTAGCACTTGGGAGCCAAGGTGGGTGGATCAC CGAGGTCAAGGAGTTGGAGACCAAGACTGGCCAACGTGGAAACATCATCTACAATACA AAAATACAAAAAATTAGCTGGCGTGGTGGCAGGCCCTGTAATCCAGCTACTTGGGAGG CTGAGGCCAGGAAATCGCTTGAACCTGGGAGGTGGAGGTTGCAAGAGGCCAGATTGCGC CATTGCACTCCAGCCTGGCGACAGAGAGACTCTGCTCTAAAAA AAAAA [A, -, T] TGGTGGCTCACGCCGTAACTCTAGCACTTGGGAGCCAAGGTGGGTGGATCACCGAG GTCAGGAGTTGGAGACCAAGACTGCCAACGTGGAAAACATCATCTACAATACAAAAA TACAAAATTAGCTGGCGTGGTGGCAGGCCCTGTAATCCAGCTACTTGGGAGGCTGA GGCACGAGAACATCGCTTGAACCTGGGAGGTGGAGGTTGCAAGAGGCCAGATTGCCATT GCACTCCAGCCTGGCGACAGAGAGACTCTGCTCTAAAAA [A, -, T] TGGGCTGTGAGGTATGCAGGGATTGATTTGGTGGGTGGATCTGCTTCTGGGATGAT GTGGATGCCCTCCCGTGGAGAGGGGAAGGGTTGATGAAAGTCCCAGGGACCTGAAAGTGTGT TCTGCAGCAATCCCCCTCCAGCAGAGATCGTGGAGGAGGAGTACAGACCGTGGAGGAT GGGGTGTGACATCCACTTGTAAATAGATGGTGGAGTCTCCACAGCTGGCACCCAGAGCT CCCCACTGAGGGCTGGGGAGCTGGGAGTATCAGGGAAATGGTGTCTTATCCAAAT |
| 18364 | TGGTGGCTCACGCCGTAACTCTAGCACTTGGGAGCCAAGGTGGGTGGATCACCGAG GTCAGGAGTTGGAGACCAAGACTGCCAACGTGGAAAACATCATCTACAATACAAAAA TACAAAATTAGCTGGCGTGGTGGCAGGCCCTGTAATCCAGCTACTTGGGAGGCTGA GGCACGAGAACATCGCTTGAACCTGGGAGGTGGAGGTTGCAAGAGGCCAGATTGCCATT GCACTCCAGCCTGGCGACAGAGAGACTCTGCTCTAAAAA [A, -, T] TGGGCTGTGAGGTATGCAGGGATTGATTTGGTGGGTGGATCTGCTTCTGGGATGAT GTGGATGCCCTCCCGTGGAGAGGGGAAGGGTTGATGAAAGTCCCAGGGACCTGAAAGTGTGT TCTGCAGCAATCCCCCTCCAGCAGAGATCGTGGAGGAGGAGTACAGACCGTGGAGGAT GGGGTGTGACATCCACTTGTAAATAGATGGTGGAGTCTCCACAGCTGGCACCCAGAGCT CCCCACTGAGGGCTGGGGAGCTGGGAGTATCAGGGAAATGGTGTCTTATCCAAAT |
| 18861 | ACTTGTAATAGATGGTGGAGTCCTCCACAGCTGGCACCAGAGCTCCCACTGAGGGCTGG GGGGGAGCTGGGAGTATCAGGGAAATGGGTGCTTTATCCAATGGCTCAAGCCAGGTG GGCTACTACCTGTTAGGGGGGTGTCTCTCTCACAAACCTGTTCTCTCCAGCT GTGGTTGATCAAGGACTCATCTGCCCTGGAGAAAATACTCAACCAAGAGCAGGGAGC CTGGGGTGTGCGGGCAGGGAGCTGGGAGTGGGGATGGGGATGAGGGTGGCATGAGCT [G, A] AGGGCAGGGCCAGGGCTGGTGTCCCTAAGGTTGATCAGACTCTGTAATATTGATTT TCCAGATGGAATAAAAGGCCGTGTAATTAACTTCACCATCAGGCCCTGAAATCCGG GGGGTAGGGGAGGGTACTTACAGGATGACAATCTGGGAGCTAGAAACTTGTAGCC AGAGAAAATCTGGGAGGTCTGAAATCTCATGTTGCTGGAGCTTGGGAAAGAGAAATCTTAG AAGCAGAAAACCTTGTAAACATAAGAATCTGGGAGGGTCTAGGATCTGAGGAGACCA |
| 20443 | TGTTGTGAAGAGCAGCTCGCTCTGTGCCGCCCTGCCCTCTGTGCTGCCCTCCATCCCTGCA GCCCACTGCGTTCTCTGGCTCCACTACCCCTCCAGTTCCAGTCTGGCCCTCT |

FIGURE 3, page 11 of 12

CCTGGTGTGTGTGTGTGTGTGTGTGTGTATGCATGCATGCATATGTGTGTCC
AGGTCTGCCCTGCCCGGATGTGACAAGTAGCGGTCTTCATGGTTGCATGTGTCTGAATT
GGTGTCTGAGCTTCACATTGTATGCGCCTGTGTGCATGTGTGTGCATGGACATGCATGCT
[G, A]

TATCTGCTGTGTTCCCTCCCCATGTGTCCCCACTGGCCTTGACATGGGAGAAGGG
CATGTGCTCAGCATATCACTCAACTGTCCACATTGGGTGGGTACCTGTGTGGTGTG
TGTGTGGGGGTGTGTCTTGAAGTGGCAGGTCCAAATGCTTAGGCAATCTGAACCTTGG
ACCTTGCAAGAGAGGAGAGATGTCCTGTAGGTGGGAGGGACAGGGAGATGCAGCAGCTGC
CCGGTGACCTTTCTGCCCTTGATGGCAAAGCTGGGGTAGGGAAAGGGAGACAAGTGCT

20881

TTGAAGTGGCAGGTCCAAATGCTTAGGCAATCTGAACCTTGGACCTTGAGAGAGGAGA
GATGTCCCTGTAGGTGGGAGGGACAGGGAGATGCAGCAGCTGCCCGGTGACCTTTCTGC
CCTTGATGGCAAAGCTGGGGTAGGGAAAGGGAGACAAGTGCTCATACTAACCTCCCTCC
CTGCCCAGGCTCCTCTGTAAGGTCTGAGTCTGTCTGTGAGCATTGCATCTGTCTGT
CTATGCCCTGATGCCCTGGATGGACAAGGGGTGTGTGTGTGTGTGTGTGTGTGTG
[A, T]
GTGTGAGGCTGCAGGAAGAGGAACAGTGGGGATGGCAGGAAAGTGGCTGTGGGTCA
GGGAGGCGAT

FIGURE 3, page 12 of 12